GREEDY APPROXIMATIONS BY SIGNED HARMONIC SUMS AND THE THUE–MORSE SEQUENCE

SANDRO BETTIN, GIUSEPPE MOLTENI, AND CARLO SANNA

Abstract. Given a real number τ, we study the approximation of τ by signed harmonic sums $\sigma_N(\tau) := \sum_{n\le N} s_n(\tau)/n$, where the sequence of signs $(s_N(\tau))_{N\in\mathbb{N}}$ is defined "greedily" by setting $s_{N+1}(\tau) := +1$ if $\sigma_N(\tau) \le \tau$, and $s_{N+1}(\tau) := -1$ otherwise. More precisely, we compute the limit points and the decay rate of the sequence $(\sigma_N(\tau) - \tau)_{N\in\mathbb{N}}$. Moreover, we give an accurate description of the behavior of the sequence of signs $(s_N(\tau))_{N\in\mathbb{N}}$, highlighting a surprising connection with the Thue–Morse sequence.

Adv. Math. **366**, 107068–42 (2020).
DOI: https://doi.org/10.1016/j.aim.2020.107068

1. Introduction

Riemann's rearrangement theorem [23, §1.1] asserts that any conditionally convergent series can be rearranged to converge to any given $\tau \in \mathbb{R} \cup \{\pm\infty\}$. The classical proof of this result is constructive: assuming $\tau \in \mathbb{R}$, one first sums the positive values until they exceed τ, then one sums the first few negatives values until going below τ, and so forth. Much in the same way, one can show that for all $\tau \in \mathbb{R}$ there exist sequences $(s_n)_{n\in\mathbb{N}}$ with $s_n \in \{\pm 1\}$ such that

$$\sum_{n=1}^{\infty} \frac{s_n}{n} = \tau.$$

A natural question is then to determine whether one can find sequences $(s_n)_{n\in\mathbb{N}}$ such that the partial sums $\sigma_n := \sum_{m=1}^{n} \frac{s_m}{m}$ converge to τ particularly quickly. The rate of convergence of σ_n to τ, however, has a clear limitation, since $|\sigma_n - \sigma_{n-1}| = \frac{1}{n}$ for all $n \in \mathbb{N}$, and so $|\sigma_{n-1} - \tau| \ge \frac{1}{2n}$ for infinitely many integers n. One can then modify the problem along two different routes. The first approach is to allow the sequence to change and study the asymptotic behavior of

$$\mathfrak{m}_N(\tau) := \min\{|\sigma_N - \tau| : s_1, \dots, s_N \in \{\pm 1\}\}.$$

This approach was considered by the authors [11], where it is shown that

$$\mathfrak{m}_N(\tau) < K_{\tau,\varepsilon} \exp\left(-\frac{1}{\log 4 - \varepsilon}(\log N)^2\right) \tag{1.1}$$

for every $\varepsilon > 0$ and some constant $K_{\tau,\varepsilon} > 0$. This inequality was obtained by interpreting the problem probabilistically, studying the rate of convergence in distribution of the random variable $\sum_{n=1}^{\infty} \frac{Z_n}{n}$, where the Z_n are independent uniformly distributed random variables in $\{-1, +1\}$.

The sequence of signs realizing the minimums $\mathfrak{m}_N(\tau)$ and $\mathfrak{m}_{N+1}(\tau)$ are not related, in general. In particular, the first one is not a subsequence of the second one, and a universal sequence $(s_n)_{n\in\mathbb{N}}$ giving the best approximation does not exist. Rather than taking the minimum over all possible choices for the $(s_n)_{n\in\mathbb{N}}$, the second approach is then to fix a specific sequence

Date: June 11, 2020.
2010 *Mathematics Subject Classification.* Primary 11J25, Secondary 11B99.
Key words and phrases. diophantine approximation; Egyptian fractions; greedy algorithms.

$(s_n)_{n\in\mathbb{N}}$ and to determine whether $|\sigma_n - \tau|$ can go to zero extremely quickly with n running along a subsequence $(n_k)_{k\in\mathbb{N}}$. A natural choice for a sequence $\big(s_n(\tau)\big)_{n\in\mathbb{N}}$ with

$$\sigma_n(\tau) := \sum_{m=1}^{n} \frac{s_m(\tau)}{m}$$

(and $\sigma_0(\tau) := 0$) converging to τ is the following:

$$s_n(\tau) := \begin{cases} 1, & \text{if } \tau \geq \sigma_{n-1}(\tau); \\ -1, & \text{if } \tau < \sigma_{n-1}(\tau). \end{cases} \tag{1.2}$$

In other words, analogously to the proof of Riemann's rearrangement theorem, one chooses $s_n(\tau) = -1$ if the partial sum $\sigma_{n-1}(\tau)$ exceeds τ, and $s_n(\tau) = 1$ otherwise (see Section 2 below for other possible choices when $\sigma_{n-1}(\tau) = \tau$). We call this a *greedy approximation to τ by signed harmonic sums* since at every step, given $\sigma_{n-1}(\tau)$, the value of $s_n(\tau)$ is chosen so that the distance of $\sigma_n(\tau)$ from τ is minimized.

In this paper we consider the second approach, analyzing the decay of $\sigma_n(\tau) - \tau$. Thus, we stress that from now on, with $s_n(\tau)$ we mean the deterministic sequence defined by (1.2). Surprisingly, the behavior of the sequence $s_n(\tau)$ is not chaotic, but is extremely structured and allows one to prove precise results on the asymptotic behavior of $\sigma_n(\tau) - \tau$.

Our first result determines the set $S'_k(\tau)$ of limit points of the sequence

$$S_k(\tau) := \big((\sigma_n(\tau) - \tau) \cdot n^k\big)_{n\in\mathbb{N}}$$

for all integers $k \in \mathbb{N}$ (the case $k = 0$ being trivial, since the sequence tends to 0). Surprisingly, only a few limit points can arise.

Theorem 1.1. *There exist sets $X_1, X_2, \dots$ which are non-empty, pairwise disjoint, countable and with $X_h \cap \overline{\mathbb{Q}} = \emptyset$ for all h, such that for all integers $k \in \mathbb{Z}_{\geq 0}$ and all $\tau \in \mathbb{R}$ one has*

$$S'_{k+1}(\tau) = \begin{cases} \{0, \pm c_0\}, & \text{if } k = 0 \text{ and } \tau \notin X_1; \\ \{\pm c_0/2\}, & \text{if } k = 0 \text{ and } \tau \in X_1; \\ \{0, \pm c_k, \pm\infty\}, & \text{if } k \geq 1 \text{ and } \tau \notin Y_{k+1}; \\ \{\pm c_k/2, \pm\infty\}, & \text{if } k \geq 1 \text{ and } \tau \in X_{k+1} = Y_{k+1} \setminus Y_k; \\ \{\pm\infty\}, & \text{if } k \geq 1 \text{ and } \tau \in Y_k; \end{cases} \tag{1.3}$$

where $Y_k := \cup_{h\leq k} X_h$ and $c_k := 2^{\binom{k}{2}} k!$.

Let $Y_\infty := \cup_{h=1}^{\infty} X_h$. Notice that Y_∞ is countable, and thus in particular has Lebesgue measure zero. From Theorem 1.1 one immediately deduces the following corollary.

Corollary 1.2. *For all integers $k \geq 0$ and all $\tau \in \mathbb{R}$ one has $S'_{k+1}(\tau) \subseteq \{0, \pm c_k, \pm c_k/2, \pm\infty\}$. Moreover, if $\tau \notin Y_\infty$ one has $S'_{k+1}(\tau) = \{0, \pm c_k, \pm\infty\}$ for all $k \geq 1$.*

One can also obtain sharp explicit bounds for $\sigma_n(\tau) - \tau$.

Corollary 1.3. *Let $k \geq 0$ and let $\tau \notin Y_k$. Then there are infinitely many n such that*

$$0 < (\sigma_n(\tau) - \tau) n^{k+1} < c_k, \tag{1.4}$$

and infinitely many n such that

$$0 < (\tau - \sigma_n(\tau)) n^{k+1} < c_k. \tag{1.5}$$

In particular, if $\tau \notin Y_\infty$, then there is a subsequence of $\big(\sigma_n(\tau) - \tau\big)_{n\in\mathbb{N}}$ that decays faster than any power of n.

Remark 1.4. Corollary 1.3 is slightly sharper than what follows from Theorem 1.1. Its proof also shows that the n satisfying one of (1.4) and (1.5) form an arithmetic progression modulo 2^k when n is sufficiently large.

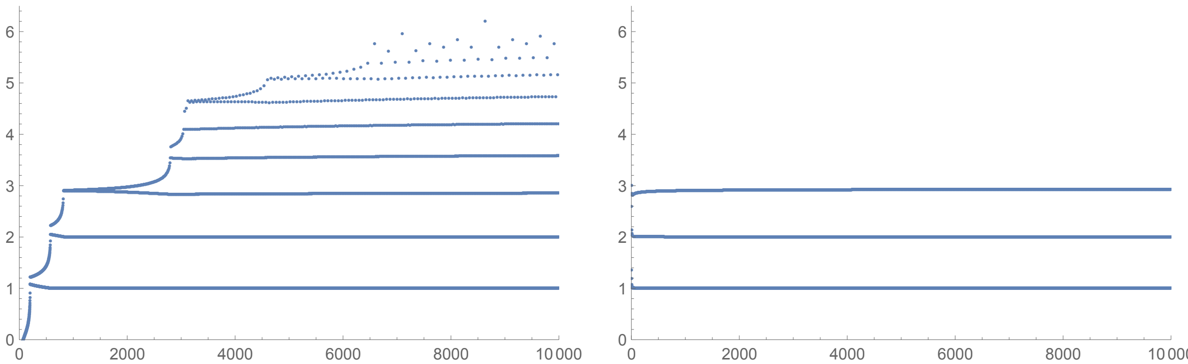

FIGURE 1. The graphs of $-\frac{\log|\sigma_n(\tau)-\tau|}{\log n}$ for $2 \le n \le 10^4$ for $\tau = \sqrt{2}+2\sqrt{5} \notin Y_\infty$ (left) and $\tau = U_{3,2} = \frac{1}{8}((2\sqrt{2}-1)\pi + \log 4) \in X_3$ (right).

For almost all τ one can even determine exactly how small $|\tau - \sigma_n(\tau)|$ can be. Indeed, one obtains that, for almost all τ, the distance $|\tau - \sigma_n(\tau)|$ is infinitely often as small as $e^{-(\log n)^2/\log 4(1+o(1))}$ and that this bound is optimal. It is quite remarkable that the sequence $s_n(\tau)$ allows one to recover, albeit for a subsequence and only for almost all τ, exactly the estimate (1.1) obtainable with probabilistic methods [11].

Theorem 1.5. *For almost all $\tau \in \mathbb{R}$ one has*

$$\liminf_{n\to\infty} \frac{\log|\tau - \sigma_n(\tau)|}{(\log n)^2} = -\frac{1}{\log 4}. \tag{1.6}$$

Remark 1.6. Equation (1.6) does not hold for all $\tau \in \mathbb{R}$. For example, by Theorem 1.1 one has that (1.6) does not hold for all $\tau \in Y_\infty$. Indeed, if $\tau \in X_k$ one has that $|\tau - \sigma_n(\tau)|$ is always larger than a constant times n^{-k}. In the opposite direction, in Proposition 5.9 below we shall show that for any function $f\colon \mathbb{N} \to \mathbb{R}_{>0}$ there exists a real number τ such that $|\tau - \sigma_n(\tau)| < f(n)$ infinitely often.

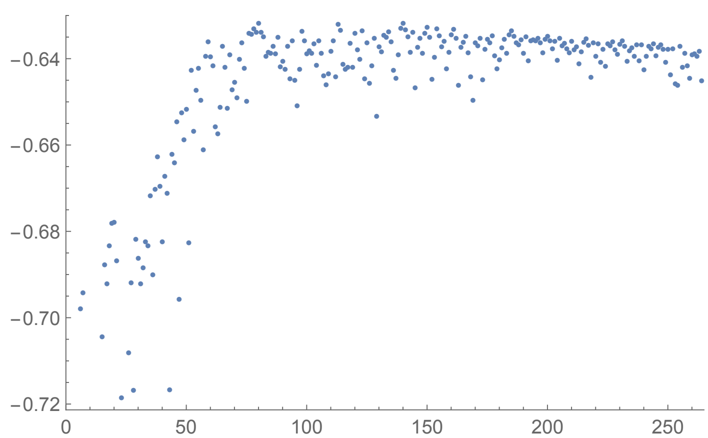

FIGURE 2. The plot of $\frac{\log|\sigma_{m_j}(\tau)-\tau|}{(\log m_j)^2}$ for $\tau = 0$ and $1 \le j \le 265$ (i.e., $m_j < 10^{10}$), where m_j is the minimal integer such that $|\sigma_{m_j}(\tau) - \tau| < |\sigma_m(\tau) - \tau|$ for all $m < m_j$. Note that the sequence is not far from $-1/\log 4 \approx -0.721\ldots$ which Theorem 1.5 predicts as the liminf for almost every τ.

The proofs of Theorem 1.1 and Theorem 1.5 are based on a surprising connection between the sequence $(s_n(\tau))_{n\in\mathbb{N}}$ and the Thue–Morse sequence $(t_n)_{n\ge 0}$. This is a binary sequence whose first few values are

$$0,1,1,0,1,0,0,1,1,0,0,1,0,1,1,0,1,0,0,1,\dots$$

and which can be defined in several equivalent ways. For example, it can be defined by setting $t_n := 0$ (respectively, $t_n := 1$) if the binary expansion of n has an even (respectively, odd) number of 1s. Alternatively, it can be defined by the recurrence relation

$$t_0 := 0, \quad t_{2n} = t_n, \quad t_{2n+1} = 1 - t_n,$$

or by the L-system which starts with 0 and at each step substitutes the digits 0 and 1 with $0,1$ and $1,0$, respectively (see [13, 14, 15, 16] and [25, Section 2.2]). Furthermore, the Thue–Morse sequence is 2-automatic, meaning that t_n is obtained by feeding a deterministic finite automaton with an output function with the base-2 representation of n [5, Example 5.1.2]. The Thue–Morse sequence has repeatedly appeared in several fields of mathematics, including dynamic systems, combinatorics, number theory and approximation theory, and has been studied extensively in its many aspects. We refer for example to [1, 12, 18, 19, 24, 26, 27, 28, 29] and to the nice survey of Allouche and Shallit [6] for a more extensive discussion of the ubiquity of the Thue–Morse sequence.

For our purposes, it is more convenient to modify the Thue–Morse sequence so that it takes its values in $\{\pm 1\}$, and thus we set $\varepsilon_n := (-1)^{t_n}$ for all $n \geq 0$.

The first connection between the Thue–Morse and our greedy sequence is apparent from the following identity [30, 35] (see [4, 7, 32] for some generalizations)

$$\prod_{n=0}^{\infty} \left(1 - \frac{1}{2n+2}\right)^{\varepsilon_n} = \frac{1}{\sqrt{2}}.$$

Shallit observed, and Allouche and Cohen [2] later proved, that for all n one has $\prod_{n=0}^{j-1}(1 - \frac{1}{2n+2})^{\varepsilon_n} > 1/\sqrt{2}$ if and only if $\varepsilon_j = 1$. In other words, passing to the logarithms, $\sum_{n=0}^{\infty} \varepsilon_n f(n)$ is the greedy series for $-\frac{1}{2}\log 2$ with respect to the weight function $f(x) := \log(1 - \frac{1}{2x+2})$.

It turns out that the connection between the sequence of signs $s_n(\tau)$ and the Thue–Morse sequence is much broader than the result of Allouche and Cohen might suggest. Indeed, $s_n(\tau)$ can be written in terms of the Thue–Morse sequence for all τ. We describe this connection in the following result, which is the key ingredient in the proof of Theorem 1.1.
We let B_r denote the block $(\varepsilon_n)_{0 \leq n < 2^r}$ for all $r \in \mathbb{Z}_{\geq 0}$. Also, given two or more vectors $v_i = (v_{i,j})_j$, with $(v_1, v_2, \dots)$ we mean the vector (or infinite sequence) obtained by concatenating the vectors $v_1, v_2, \dots$ Moreover, given a vector v and a scalar κ, with $\kappa \cdot v$ we mean the vector v multiplied by the scalar κ; for example, $(-1) \cdot B_r = (-\varepsilon_n)_{0 \leq n < 2^r}$.

Theorem 1.7. *Let $\tau \in \mathbb{R}$. Then there exist a non-decreasing sequence $(k_i)_{i \in \mathbb{N}}$, with $k_i \in \mathbb{Z}_{\geq 0}$ for all i, and a sequence $(\kappa_i)_{i \in \mathbb{N}}$, with $\kappa_i \in \{\pm 1\}$, such that*

$$(s_n(\tau))_{n \in \mathbb{N}} = (\kappa_1 \cdot B_{k_1}, \kappa_2 \cdot B_{k_2}, \kappa_3 \cdot B_{k_3}, \dots). \tag{1.7}$$

Moreover, $\lim_{i \to \infty} k_i = +\infty$ if $\tau \notin Y_\infty$, whereas if $\tau \in X_k$ for some k, then $k_i = k$ and $\kappa_i = 1$ for i large enough.

Since for all $r_1, r_2 \in \mathbb{Z}_{\geq 0}$ with $r_1 \leq r_2$ the block B_{r_2} contains the block B_{r_1}, we immediately deduce the following corollary.

Corollary 1.8. *If $\tau \notin Y_\infty$, then for all $r \geq 0$ the sequence $(s_n(\tau))_{n \in \mathbb{N}}$ contains the block B_r infinitely many times.*

We remark that Theorem 1.7 characterizes the elements of the exceptional sets X_k as the real numbers such that the expansion (1.7) is eventually periodic with repeating block B_k (in fact, this will actually be our definition of X_k). In particular, the elements in X_k can be written as a rational number plus $U_{k,m}$ for some $m \in [0, 2^k)$, where

$$U_{k,m} := \sum_{n=1}^{+\infty} \frac{f_k(n+m)}{n}, \tag{1.8}$$

with f_k being the 2^k-periodic function such that $f_k(n) = \varepsilon_n$ for all $n \in [0, 2^k)$. (Notice that $U_{k,m} = -U_{k,m+2^{k-1}}$ for all $m \in [0, 2^{k-1})$.) For small k one can easily write the values $U_{k,m}$ explicitly. For example,

$$U_{1,0} = -\log 2, \qquad U_{2,0} = -\tfrac{1}{4}(\pi + \log 4), \qquad U_{2,1} = -\tfrac{1}{4}(\pi - \log 4). \tag{1.9}$$

and so, in particular,

$$X_1 \subseteq \mathbb{Q} \pm \log 2, \qquad X_2 \subseteq \mathbb{Q} \pm \tfrac{1}{4}(\pm\pi \pm \log 4).$$

In general the constants $U_{k,m}$ can be written in terms of logarithms of cyclotomic units in $\mathbb{Q}(\xi)$, where ξ is a primitive 2^k-root of unity. Baker's theorem can then be used to show that $U_{k,m}$ is transcendental for all choices of k and m. We refer to Section 4 for more details on the sets X_k and the constants $U_{k,m}$.

Since the Thue–Morse sequence plays a special role in our work, we examine the case of the constant

$$\tau_0 := \sum_{n=1}^{+\infty} \frac{\varepsilon_{n-1}}{n},$$

in more detail. This constant appears as the value at $s = 1$ of the Thue-Morse Dirichlet series $\sum_{n=1}^{\infty} \frac{\varepsilon_{n-1}}{n^s}$, which was studied in [2, 3]. In Section 6 we prove that this series converges, and that the sequence $(\varepsilon_n)_{n\geq 1}$ indeed coincides with the sequence of signs produced by the greedy algorithm (much as in the aforementioned work of Allouche and Cohen [2]). In other words, $\varepsilon_{n-1} = s_n(\tau_0)$ for all $n \in \mathbb{N}$. We remark that this is not immediate. Since the Thue–Morse sequence is not eventually periodic, then $\tau_0 \notin Y_\infty$ and the results in Corollaries 1.2–1.3 apply to τ_0. However, since we know the sequence $(s_n(\tau_0))_{n\in\mathbb{N}}$ exactly, we are able to prove these results in an explicit form.

Theorem 1.9. *For all $n \geq 1$ we have $s_n(\tau_0) = \varepsilon_{n-1}$. Moreover, let $k \geq 1$, and let $n = 2^k n'$ with n' odd. Then $|\sigma_n - \tau_0| n^k \leq c_{k-1}$, and $(\tau_0 - \sigma_n(\tau_0)) n^{k+1} \sim_k \varepsilon_n c_k$ as n' goes to infinity. Finally,* (1.6) *holds for $\tau = \tau_0$ in the more precise form*

$$\liminf_{n\to\infty} \frac{\log|\tau_0 - \sigma_n(\tau_0)| + \frac{1}{\log 4}(\log n)^2}{\log n \log\log n} = \frac{1}{\log 2}. \tag{1.10}$$

The first part of this theorem allows one to compute τ_0 very quickly. The decimal representation of τ_0 with 50 correct digits is:

$$\tau_0 = 0.39876108810841881240743054440027306033680891546719\ldots.$$

In this paper we focused on the case of signed harmonic sums, but our work could also be extended with little effort to signed sums of the form $\sum_n s_n^\alpha(\tau)\, n^{-\alpha}$ for any $\alpha \in (0, 1]$ (and, to some extent, also to more generic weight functions). In particular, Theorems 1.1, 1.5 and 1.7 still hold (removing the statement $X_h \cap \overline{\mathbb{Q}} = \emptyset$), once one replaces $S_k(\tau)$ with $S_k^\alpha(\tau) := \big((\sigma_n(\tau) - \tau) \cdot n^{k+\alpha-1}\big)_{n\in\mathbb{N}}$ and c_k with $c_k^\alpha := 2^{\binom{k}{2}} \alpha(\alpha+1)\cdots(\alpha+k-1)$.

One can also generalize the problem to other directions. For example, one can consider the analogous problems in higher dimensions or can require that s_n takes its values in the k-th roots of unity rather than in $\{\pm 1\}$. For $k \neq 2$ the greedy algorithm again produces a representation for every complex τ. The rate of convergence has a similar behavior, but some new phenomena appear, and it is possible that in this case the role of the Thue–Morse sequence is played by its analogue for the base-k representations of numbers. We still need to study these generalizations in depth.

Acknowledgements. S. Bettin is member of the INdAM group GNAMPA. G. Molteni and C. Sanna are members of the INdAM group GNSAGA. The work of the first and second author is partially supported by PRIN 2015 “Number Theory and Arithmetic Geometry”. C. Sanna is supported by a postdoctoral fellowship of INdAM. Part of this work was done during a visit, partially supported by INdAM, of the first author to the Centre de Recherches Mathématiques in Montréal. He would like to thank this institution for its hospitality. He would also like to thank M. Radziwiłł for several useful discussions as well as for the help with the C code used for Figure 2. The authors would like to thank the anonymous referees for carefully reading the paper and for giving many suggestions that improved its quality.

Notation. For ease of notation, in the following we shall usually omit the dependence of τ in $s_n(\tau)$ and $\sigma_n(\tau)$. We stress, however, that the sequences $(\sigma_n)_n$ and $(s_n)_n$ are the sequences obtained from the greedy algorithm (1.2) applied to τ. Finally, given a set $I \subseteq \mathbb{R}$, we indicate with $\chi_I(x)$ the characteristic function of I.

2. Preliminary considerations and the proof of Theorem 1.1

In the definition (1.2) of the sequence $(s_n)_n$ we defined $s_n := +1$ whenever $\sigma_{n-1} = \tau$. Of course, other choices are also possible, and one could even decide to stop the algorithm whenever such equality is achieved. Another natural choice would be that of defining $(s_n)_n$ as in (1.2) when $\tau \geq 0$ and putting instead

$$s_n(\tau) := \begin{cases} 1, & \text{if } \tau > \sigma_{n-1}(\tau); \\ -1, & \text{if } \tau \leq \sigma_{n-1}(\tau); \end{cases}$$

when $\tau < 0$. Notice that this alternative definition would ensure that $(s_n(\tau))_n$ is odd in τ for all $\tau \neq 0$. The conclusions of this paper are essentially independent of the choice made in these special cases, so we chose the definition (1.2) since it slightly simplifies some statements. In any case, independently of the choice made, the equality $\sigma_n = \tau$ (which clearly requires $\tau \in \mathbb{Q}$) could only be achieved at most one time.

Proposition 2.1. *Let $h \in \mathbb{Z}$ and $k \in \mathbb{N}$ with $(h,k) = 1$. Let $(r_n)_n$ be a sequence taking values in $\{\pm 1\}$. Then there exists at most one $N \in \mathbb{N}$ such that $\sum_{n=1}^N r_n/n = h/k$. Moreover, if such an $N \geq 2$ exists, then it satisfies $N \leq 3\log k$.*

Proof. Suppose, to get a contradiction, that there exist $N, M \in \mathbb{N}$, with $N < M$, such that $\sum_{n=1}^N r_n/n = \sum_{n=1}^M r_n/n = \tau$. In particular,

$$\sum_{n=N+1}^{M} \frac{r_n}{n} = 0. \tag{2.1}$$

Let 2^ν be the greatest power of 2 dividing some integer in the range $N+1, \dots, M$. Notice that we can assume that $M \geq N+2$ (otherwise the sum contains a unique term, which cannot be 0), and in particular $\nu \geq 1$. Now there exists a unique m in $N+1, \dots, M$ that is divisible by 2^ν. Indeed, if there were two such integers m, m' with $m < m'$, then, letting $m = 2^\nu g$ with g odd, we would have that $N+1 \leq m < m + 2^\nu \leq m' \leq M$. But $m + 2^\nu = 2^{\nu+1}\frac{g+1}{2}$, which contradicts the maximality of ν. Writing $\ell := \operatorname{lcm}(n+1, \dots, n')$, we have that $\sum_{n=N+1}^M r_n \frac{\ell}{n}$ is odd and hence non-zero, contradicting (2.1).

To prove that $N \leq 3\log k$ one can proceed in a similar way, observing that if $\sum_{n=1}^N \frac{r_n}{n} = \frac{h}{k}$, then k is divisible by all prime powers in $(N/2, N]$. In fact, let p^ν be any prime power in $(N/2, N]$. Then $N < 2p^\nu$, so that p^ν is the unique number in $[1, N]$ that is divisible by p^ν. Hence, when we multiply the sum $\sum_{n=1}^N \frac{r_m}{m}$ by the least common multiply of the denominators, we get an integer which is congruent to r_{p^ν} modulo p^ν; in particular, it is not divisible by p. Thus $k \geq e^{\psi(N)-\psi(N/2)}$, where $\psi(x)$ is the Chebyshev's function. By [31, Theorem 8] one easily deduces that $\psi(N) - \psi(N/2) > N/3$ for all integers $N \geq 2$ and the result follows. $\square$

The set $W := \{\tau \in \mathbb{R}\colon \sigma_N = \tau \text{ for some } N \in \mathbb{N}\}$ does not seem easy to characterize, similarly to what happens in analogous problems on Egyptian fractions (cf. [21, §D11]). It is clear, however, that $W \subseteq \mathbb{Q}$, that W is nonempty (for example, it contains 1, $1/2 = 1 - 1/2$, $1/6 = 1 - 1/2 - 1/3$) and that it is a decidable set of $\mathbb{Q}$, as there is an algorithm that, for every $\tau \in \mathbb{Q}$, is able to determine if $\tau \in W$ in a finite number of steps. We shall not consider the problem of studying W in this paper. However, we remark that by Proposition 2.1 the equality $\sigma_N = \tau$ can be achieved only once for each τ, and so W does not depend on which version of the greedy algorithm we use.

2.1. **Sequence of signs and the corresponding inequalities.** A sequence of inequalities corresponds to every sequence of signs $(s_n)_n$; indeed, one has

$$s_{n+1} = +1 \iff \sigma_n \le \tau \qquad \text{and} \qquad s_{n+1} = -1 \iff \tau < \sigma_n. \tag{2.2}$$

In particular, one has that $s_{n+1} = +1$ and $s_{n'+1} = -1$ with $n' > n$ if and only if $\sigma_n \le \tau < \sigma_{n'}$, whereas the reversed inequality holds if $s_{n+1} = -s_{n'+1} = -1$. In both cases

$$s_{n+1} \cdot (\tau - \sigma_n) = |\sigma_n - \tau| \le |\sigma_{n'} - \sigma_n| = \Bigg|\sum_{m=n+1}^{n'} s_m/m\Bigg|. \tag{2.3}$$

Thus $|\sigma_n - \tau| \le (n' - n)/(n+1)$ and stronger conclusions may be drawn if one is able to bound the sum $\sum_{m=n+1}^{n'} s_m/m$ more effectively, which in turn amounts to controlling the sequence of signs $(s_m)_m$. The following proposition gives a first example of this phenomenon.

Proposition 2.2. *Let $\tau \in \mathbb{R}$. Then after the first change of sign the sequence $(s_n)_n$ has no three consecutive equal terms. As a consequence, $|\sigma_n - \tau| \le 2/(n+1)$ for all n following the first sign change. In particular, the series $\sum_{n=1}^{\infty} s_n/n$ converges to τ.*

Proof. The definition of the greedy process and the fact that the harmonic series diverges show that the sequence $(s_n)_n$ contains infinitely many changes of sign. In particular, to prove the first assertion of the proposition it suffices to show that one cannot have

$$-s_n = s_{n+1} = s_{n+2} = s_{n+3} = 1 \quad \text{or} \quad -s_n = s_{n+1} = s_{n+2} = s_{n+3} = -1$$

for n large enough. Indeed, if the former equality is satisfied, then by (2.2) we have

$$\sigma_n < \sigma_{n+1} < \sigma_{n+2} \le \tau < \sigma_{n-1},$$

whence

$$1/n = \sigma_{n-1} - \sigma_n > \sigma_{n+2} - \sigma_n = 1/(n+2) + 1/(n+1),$$

and this is impossible for $n \ge 2$. Analogously, one excludes the second case. The bound for $\sigma_n - \tau$, and thus the convergence of the series, follows immediately from (2.3). $\square$

Since $|\sigma_{n+1} - \sigma_n| = \frac{1}{n+1}$ for all n, one cannot significantly improve the inequality of Proposition 2.2 for all large enough n. However, the signs patterns exhibited by Theorem 1.7 allows us to draw stronger conclusions for arbitrarily large n. The following lemma allows one to control the sum $\sum_{m=n+1}^{n'} s_m/m$ when s_m is given by the ± 1-valued Thue–Morse sequence $\varepsilon_n := (-1)^{t_n}$. Before stating it, we define the function $g_k(x)$ for $k \in \mathbb{Z}_{\ge 0}$ and $x > 0$ as

$$g_k(x) := \sum_{0 \le \ell < 2^k} \frac{\varepsilon_\ell}{x + \ell}. \tag{2.4}$$

This function, which could also be defined as a 2^{k-1}-th iteration of the operator $(\Delta_\alpha f)(x) := f(x) - f(x + \alpha)$ on the function $g_0(x) := \frac{1}{x}$ (cf. Lemma 3.1 below), will play an important role in the proofs of our results.

Lemma 2.3. *Let $r = \sum_{0\le j\le q} 2^{h_j} > 0$ with $h_0 > h_1 > \cdots > h_q$. Then, as $x \to \infty$, we have*

$$\sum_{0\le \ell<r} \frac{\varepsilon_\ell}{x+\ell} \sim (-1)^q \frac{c_{h_q}}{x^{h_q+1}}.$$

In particular, for all $k \in \mathbb{Z}_{\ge 0}$ we have

$$g_k(x) \sim \frac{c_k}{x^{k+1}}, \tag{2.5}$$

as $x \to +\infty$. Furthermore, for $x \in \mathbb{R}_{>0}$, the function $g_k(x)$ is positive, decreasing and satisfies $g_k(x) < c_k x^{-k-1}$.

We postpone the proof of this lemma to Section 3 (cf. Corollary 3.4 and Lemma 3.5). We now show how this lemma, in combination with Theorem 1.7, implies Theorem 1.1. Before moving to the proof, we record the "folding" property of the Thue–Morse sequence in terms of ε_n:

$$\varepsilon_{n+m} = \varepsilon_n \varepsilon_m \qquad \forall n, m \ge 0 \quad \text{with} \quad n < 2^h \tag{2.6}$$

where h is the greatest nonnegative integer ν such that 2^ν divides m. In particular,

$$\varepsilon_{n+2^h} = -\varepsilon_n \qquad \forall n, m \ge 0 \quad \text{with} \quad n < 2^h \tag{2.7}$$

(see [13] and [25, Section 2.2]).

Proof of Theorem 1.1. We only deal with the case $k \ge 1$, the case $k = 0$ being very similar. Also, we postpone the proof that the sets $X_1, X_2, \ldots$ are pairwise disjoint, countable and with $X_k \cap \overline{\mathbb{Q}} = \emptyset$ for all k to Section 4.

First, assume that $\tau \notin Y_{k+1}$. By Theorem 1.7, there exist $N, j \in \mathbb{N}$ such that

$$(s_n)_{n\ge N} = (\kappa_j \cdot B_{k_j}, \kappa_{j+1} \cdot B_{k_{j+1}}, \kappa_{j+2} \cdot B_{k_{j+2}}, \ldots) \tag{2.8}$$

with $k_i \ge k+2$ for all $i \ge j$. By (2.6), for $h, k \ge 0$ we have

$$B_{k+h} = (\varepsilon_0 \cdot B_{k+2}, \varepsilon_1 \cdot B_{k+2}, \ldots, \varepsilon_h \cdot B_{k+2}).$$

Hence, (2.8) can be rewritten as

$$(s_n)_{n\ge N} = (\delta_0 \cdot B_{k+2}, \delta_1 \cdot B_{k+2}, \delta_2 \cdot B_{k+2}, \ldots)$$

where $(\delta_m)_m$ is a sequence with values in $\{\pm 1\}$.

Now let $M = N + 2^{k+2} m$ with $m \ge 0$, so that $s_{n+M} = \delta_m \varepsilon_n$ for $0 \le n < 2^{k+2}$. In particular, $s_{M+2^{k+1}} = -s_M$ and so by (2.3) and (2.5) we have

$$|\sigma_{M-1} - \tau| \le |\sigma_{M-1+2^{k+1}} - \sigma_{M-1}| = \left| \sum_{0\le \ell<2^{k+1}} \frac{\delta_m \varepsilon_\ell}{M+1+\ell} \right| \sim \frac{c_{k+1}}{M^{k+2}},$$

as $m \to \infty$ (i.e., $M \to \infty$). It follows that, writing $0 \le r < 2^{k+2}$ as in Lemma 2.3, we have

$$M^{k+1}(\sigma_{M-1+r} - \tau) = M^{k+1}(\sigma_{M-1} - \tau) + M^{k+1} \sum_{0\le \ell<r} \frac{\delta_m \varepsilon_\ell}{M+\ell} = o(1) + \frac{(-1)^q \delta_m c_{h_q}}{M^{h_q-k}}(1+o(1))$$

as $m \to \infty$, whence

$$\lim_{m\to\infty} \delta_m (N + 2^{k+2} m)^{k+1} (\sigma_{N+2^{k+2}m-1+r} - \tau) = \begin{cases} 0, & \text{if } r \in \{0, 2^{k+1}\}; \\ c_k, & \text{if } r = 2^k; \\ -c_k, & \text{if } r = 2^k + 2^{k+1}; \\ (-1)^q \infty, & \text{if } r \notin \{0, 2^k, 2^{k+1}, 2^k + 2^{k+1}\}. \end{cases}$$

The fact that $S'_{k+1}(\tau) = \{0, \pm c_k, \pm\infty\}$ then follows.

Now assume that $\tau \in X_{k+1}$. By Theorem 1.7, there exists $N \in \mathbb{N}$ such that

$$(s_n)_{n\ge N} = (B_{k+1}, B_{k+1}, B_{k+1}, \ldots).$$

Since the sets X_h are pairwise disjoint, we have that $\tau \notin Y_k$. In particular, the above argument gives

$$\lim_{m\to\infty}(N+2^{k+1}m)^{k+1}(\sigma_{N+2^{k+1}m-1+r}-\tau)=(-1)^q\infty$$

for $0<r<2^{k+1}$, $r\neq 2^k$. We shall now show that the above limit is $c_k/2$ when $r=0$. Similarly one shows that the limit is $-c_k/2$ when $r=2^k$.
By grouping the sums in the B_{k+1} blocks, one sees that

$$\tau=\sum_{n=1}^{\infty}\frac{s_n}{n}=\sigma_{N-1}+\sum_{h=0}^{\infty}\sum_{\ell=0}^{2^{k+1}-1}\frac{\varepsilon_\ell}{N+h2^{k+1}+\ell}=\sigma_{N-1}+\sum_{h=0}^{\infty}g_{k+1}(N+h2^{k+1}).$$

It follows that

$$\tau-\sigma_{N+2^{k+1}m-1}=\sum_{h=m}^{\infty}g_{k+1}(N+h2^{k+1}). \tag{2.9}$$

By Lemma 2.3 g_{k+1} is positive and decreasing. Thus by (2.5) as $x\to+\infty$ one has

$$\begin{aligned}\sum_{h\geq 0}g_{k+1}(x+h2^{k+1})&=\int_0^{+\infty}g_{k+1}(x+u2^{k+1})\,\mathrm{d}u+O(x^{-k-2})\\&=\frac{1}{2^{k+1}}\int_x^{+\infty}g_{k+1}(u)\,\mathrm{d}u+O(x^{-k-2})\\&=\frac{1}{2^{k+1}}\int_x^{+\infty}\Big(\frac{c_{k+1}}{u^{k+2}}+o(u^{-k-2})\Big)\,\mathrm{d}u+O(x^{-k-2})\\&=\frac{c_{k+1}}{2^{k+1}(k+1)x^{k+1}}+o(x^{-k-1})=\frac{c_k}{2x^{k+1}}+o(x^{-k-1}),\end{aligned}$$

since $(k+1)2^kc_k=c_{k+1}$. Thus, by (2.9) we obtain

$$\lim_{m\to\infty}(N+2^{k+1}m)^{k+1}(\sigma_{N+2^{k+1}m-1}-\tau)=c_k/2,$$

as claimed. Collecting the above results one obtains $S'_{k+1}(\tau)=\{\pm c_k/2,\pm\infty\}$ for $\tau\in X_{k+1}$.

Finally, the case $\tau\in Y_k$ of (1.3) can be proven easily along the same lines. □

Proof of the Corollary 1.3. Since $\tau\notin Y_k$, then by Theorem 1.7 one has that there exists $N\in\mathbb{N}$ such that the sequence $(s_n)_{n\geq N}$ can be written as a concatenation of blocks $\pm B_{k+1}=\pm(B_k,-B_k)$. It follows that the sequence $(s_n)_n$ contains infinitely many blocks $(B_k,-B_k)$ and infinitely many blocks $(-B_k,B_k)$. Now if $(s_r)_{m\leq r<m+2^{k+1}}=(B_k,-B_k)$ then applying (2.3) with $n=m-1$ and $n'=m+2^k-1$, we obtain

$$0\leq(\tau-\sigma_{m-1})\leq\left|\sum_{r=m}^{m+2^k-1}\varepsilon_r/r\right|=g_k(m)<c_km^{-k-1}<c_k(m-1)^{-k-1},$$

where in the third inequality we used Lemma 2.3. Also, by Proposition 2.1, the first inequality is strict if m is large enough. It follows that (1.5) is satisfied for infinitely many m and one proves in the same way that the same holds for (1.4). □

3. Thue–Morse sums and the proof of Theorem 1.7

3.1. **The function $g_k(x)$ and other Thue–Morse sums.** We recall that the function $g_k(x)$ was defined in (2.4) as a certain logarithmic average of the Thue–Morse sequence, and that we defined the operator $\Delta_\alpha f$ as $(\Delta_\alpha f)(x):=f(x)-f(x+\alpha)$.

Lemma 3.1. *For all $k\in\mathbb{N}$ and $x>0$ we have*

$$g_k(x)=(\Delta_{2^{k-1}}\circ\cdots\circ\Delta_2\circ\Delta_1)g_0(x) \tag{3.1}$$

or, equivalently,

$$g_k(x) = (-1)^k \int_0^{2^{k-2}} \cdots \int_0^{2^0} g_0^{(k)}(x + u_1 + \cdots + u_k)\,\mathrm{d}u_k \cdots \mathrm{d}u_2\,\mathrm{d}u_1. \tag{3.2}$$

Proof. Since g_0 admits derivatives of any order for $x > 0$, Equation (3.2) follows immediately from (3.1). Now we prove (3.1) by induction on k. For $k = 1$, the equality is immediate. Now let $k \geq 2$ and suppose the claim is true for $k-1$. Then we split the range of summation $[0, 2^k)$ in the definition of g_k into $[0, 2^{k-1}) \cup [2^{k-1}, 2^k)$ and shift the variable in the second range by 2^{k-1}. We get

$$g_k(x) = \sum_{0\leq \ell < 2^k} \frac{\varepsilon_\ell}{x+\ell} = \sum_{0\leq \ell < 2^{k-1}} \frac{\varepsilon_\ell}{x+\ell} + \sum_{0\leq \ell < 2^{k-1}} \frac{\varepsilon_{\ell+2^{k-1}}}{x+2^{k-1}+\ell} = g_{k-1}(x) - g_{k-1}(x+2^{k-1}),$$

since $\varepsilon_{\ell+2^{h-1}} = -\varepsilon_\ell$ by the folding property (2.7). Thus, by the inductive hypothesis we have

$$g_k(x) = \Delta_{2^{k-1}} g_{k-1}(x) = (\Delta_{2^{k-1}} \circ \cdots \circ \Delta_2 \circ \Delta_1) g_0(x),$$

as claimed. □

The following lemma shows that any logarithmic average of the Thue–Morse sequence can be expressed as a combination of the functions $g_k(x)$.

Lemma 3.2. *Let $r := \sum_{0\leq j\leq q} 2^{h_j} > 0$ with $h_0 > h_1 > \cdots > h_q$, and let $r_j := \sum_{i<j} 2^{h_i}$. Then*

$$\sum_{0\leq \ell < r} \frac{\varepsilon_\ell}{x+\ell} = \sum_{0\leq j\leq q} (-1)^j g_{h_j}(x+r_j).$$

Proof. The proof is by induction on q. The case $q = 0$ holds true by definition. Let $q \geq 1$, and suppose the claim is true for $q-1$. The definition of r shows that $r = r_q + 2^{h_q}$, so that, splitting the range of summation $[0, r)$ into $[0, r_q) \cup [r_q, r_q + 2^{h_q})$ and shifting the variable in the second range by r_q, we get

$$\sum_{0\leq \ell < r} \frac{\varepsilon_\ell}{x+\ell} = \sum_{0\leq \ell < r_q} \frac{\varepsilon_\ell}{x+\ell} + \sum_{0\leq \ell < 2^{h_q}} \frac{\varepsilon_{\ell+r_q}}{x+r_q+\ell}.$$

The dyadic representation of r_q contains $q-1$ nonzero digits, hence the inductive hypothesis may be applied to the first term. Moreover, in the second sum $\ell < 2^{h_q} < 2^{h_{q-1}} \leq r_q$, thus $\varepsilon_{\ell+r_q} = \varepsilon_\ell \varepsilon_{r_q} = (-1)^q \varepsilon_\ell$, so that

$$\sum_{0\leq \ell < r} \frac{\varepsilon_\ell}{x+\ell} = \sum_{0\leq j\leq q-1} (-1)^j g_{h_j}(x+r_j) + (-1)^q \sum_{0\leq \ell < 2^{h_q}} \frac{\varepsilon_\ell}{x+r_q+\ell}.$$

By definition the second sum on the right is $g_{h_q}(x+r_q)$ and the proof is complete. □

We now describe the asymptotic behavior of $g_k(x)$.

Lemma 3.3. *Let $k \geq 0$. Then, as $x \to \infty$, we have $g_k(x) = \frac{c_k}{x^{k+1}} + O_k(x^{-k-2})$, where c_k is as in Theorem 1.1.*

Proof. We show that $g_k(x) = c_k/x^{k+1} + \phi_k(1/x)$ with ϕ_k analytic in a neighborhood of 0 and with a zero of order at least $k+2$ at $x = 0$. This is obvious for $k = 0$. Suppose it is true for $k \geq 0$. We have

$$g_{k+1}(x) = \frac{c_k}{x^{k+1}} + \phi_k(1/x) - \frac{c_k}{(x+2^k)^{k+1}} - \phi_k(1/(x+2^k)) = \frac{(k+1)2^k c_k}{x^{k+2}} + \phi_{k+1}(1/x)$$

with

$$\phi_{k+1}(x) = \phi_k(x) - \phi_k(x(1+2^k x)^{-1}) - c_k x^{k+1} \left((1+2^k x)^{-k-1} - 1 + (k+1)2^k x \right).$$

Clearly $\phi_{k+1}(x)$ is analytic in a neighborhood of 0 and has a zero of order at least $k+3$ at $x = 0$. Moreover, by definition, $(k+1)2^k c_k = 2^{\frac{k(k-1)}{2}+k}(k+1)! = c_{k+1}$, as desired. □

Corollary 3.4. *With the same notation as in Lemma 3.2, as $x \to \infty$ we have*

$$\sum_{0 \le \ell < r} \frac{\varepsilon_\ell}{x+\ell} \sim (-1)^q \frac{c_{h_q}}{x^{h_q+1}}.$$

3.2. **Inequalities for $g_k(x)$.**

Lemma 3.5. *For all $m, k \ge 0$ and all $x > 0$ we have*

$$2^{\binom{k}{2}} \frac{(m+k)!}{(x+2^k)^{m+k+1}} < (-1)^m g_k^{(m)}(x) < 2^{\binom{k}{2}} \frac{(m+k)!}{x^{m+k+1}}.$$

In particular, for all $k \ge 0$, $g_k(x)$ is positive and decreasing for $x > 0$.

Proof. The claim is clear for $k = 0$. Suppose $k \ge 1$. The operator Δ_α commutes with the differentiation. Thus (3.1) and (3.2) applied to $g_k^{(m)}$ show that

$$(-1)^m g_k^{(m)}(x) = \int_0^{2^{k-1}} \int_0^{2^{k-2}} \cdots \int_0^{2^0} (-1)^{m+k} g_0^{(m+k)}(x + u_1 + \cdots + u_k)\,\mathrm{d}u_k \cdots \mathrm{d}u_2\,\mathrm{d}u_1.$$

The measure of the domain of integration is $2^{\binom{k}{2}}$ and so the claim follows, since for $x < u < x + 2^{k-1} + \cdots + 2^0 = x + 2^k - 1$ one has

$$\frac{(m+k)!}{(x+2^k)^{m+k+1}} < (-1)^{m+k} g_0^{(m+k)}(u) = \frac{(m+k)!}{u^{m+k+1}} < \frac{(m+k)!}{x^{m+k+1}}. \qquad \square$$

Corollary 3.6. *For all $k \ge 0$ and all $x > 0$ we have*

$$g_k(x) < g_{k-1}(x).$$

Proof. We have $g_k(x) = g_{k-1}(x) - g_{k-1}(x + 2^{k-1}) < g_{k-1}(x)$. $\square$

Lemma 3.7. *Let $k \ge 0$. Then for all $x \ge (k+1)2^{k+2}$ we have $g_k(x) < \frac{4}{3} g_k(x + 2^k)$.*

Proof. For $a, k \ge 0$, let

$$h_k(x, a) := 4 g_k(x + a) - 3 g_k(x).$$

The operator Δ_α is linear, and thus for all $k \ge 1$ we have

$$h_k(x, a) = h_{k-1}(x, a) - h_{k-1}(x + 2^{k-1}, a).$$

In particular, the analogues of the identities (3.1) and (3.2) also hold for h_k. Moreover, for $m \ge 0$ and $a/x < \sqrt[m+1]{4/3} - 1$ we have

$$(-1)^m h_0^{(m)}(x, a) = m! \left(\frac{4}{(x+a)^{m+1}} - \frac{3}{x^{m+1}} \right) > 0.$$

Reasoning as in the proof of Lemma 3.5 we get that

$$h_k(x, a) > 0$$

for $a/x < \sqrt[k+1]{4/3} - 1$. Now for $x \ge 0$ and $0 \le \rho \le 1$ we have $(1+x)^\rho - 1 \ge \rho x + \rho(\rho - 1)\frac{x^2}{2}$ and so $\sqrt[k+1]{4/3} - 1 \ge \frac{5k+6}{18(k+1)^2} > \frac{1}{4(k+1)}$. The lemma then follows by taking $a = 2^k$. $\square$

Lemma 3.8. *For $x \ge 2^{k+1}k$ and $0 \le h < k$ we have $g_k(x) < \frac{1}{2} g_h(x + 2^k)$.*

Proof. By Corollary 3.6, it suffices to prove the claim when $h = k - 1$.
Applying Lemma 3.7 twice, for $x \ge 2^{k+1}k$ we have

$$\begin{aligned} g_k(x) - \tfrac{1}{2} g_{k-1}(x + 2^k) &= g_{k-1}(x) - g_{k-1}(x + 2^{k-1}) - \tfrac{1}{2} g_{k-1}(x + 2^k) \\ &< \tfrac{1}{3} g_{k-1}(x + 2^{k-1}) - \tfrac{1}{2} g_{k-1}(x + 2^k) < -\tfrac{1}{18} g_{k-1}(x + 2^k) < 0, \end{aligned}$$

as desired. $\square$

Lemma 3.9. *Let $r = \sum_{0\le j\le q} 2^{h_j} > 0$ with $h_0 > h_1 > \cdots > h_q$. Then, if $r < 2^k$ and $x \ge 2^{k+1}(k+1)$, we have*

$$(-1)^q \sum_{0\le \ell<r} \frac{\varepsilon_\ell}{x+\ell} > g_k(x-2^k) > g_k(x) > 0.$$

Proof. First we observe that the inequality $g_k(x-2^k) > g_k(x) > 0$ follows by Lemma 3.5. Write $r_j = \sum_{i<j} 2^{h_i}$; by Lemma 3.2 we have

$$(-1)^q \sum_{0\le \ell<r} \frac{\varepsilon_\ell}{x+\ell} - g_k(x-2^k) = \sum_{0\le j\le q} (-1)^{j+q} g_{h_j}(x+r_j) - g_k(x-2^k).$$

If q is odd this is

$$\big(g_{h_q}(x+r_q) - g_{h_{q-1}}(x+r_{q-1})\big) + \cdots + \big(g_{h_3}(x+r_3) - g_{h_2}(x+r_2)\big) \\ + \big(g_{h_1}(x+r_1) - g_{h_0}(x+r_0) - g_k(x-2^k)\big).$$

For all $\ell \ge 1$ we have $r_\ell = r_{\ell-1} + 2^{h_{\ell-1}}$ and thus, by Lemma 3.8, for all $x \ge 2^{k+1}k$ we have

$$g_{h_\ell}(x+r_\ell) - g_{h_{\ell-1}}(x+r_{\ell-1}) > \tfrac{1}{2} g_{h_\ell}(x+r_\ell) > 0.$$

Moreover, since $r_0 = 0$ and $h_0 < k$, by Lemma 3.8 for $x > 2^{k+1}(k+1) > 2^{k+1}k + 2^k$ we have

$$g_{h_1}(x+r_1) - g_{h_0}(x+r_0) - g_k(x-2^k) > g_{h_1}(x+r_1) - \tfrac{3}{2} g_{h_0}(x+r_0) > 0.$$

The case q even is analogous and slightly simpler. □

Corollary 3.10. *Let $0 < r < 2^k$ and $x \ge 2^{k+1}(k+1)$. Then*

$$\operatorname{sgn}\Bigg(\sum_{0\le \ell<r} \frac{\varepsilon_\ell}{x+\ell}\Bigg) = -\varepsilon_r.$$

Proof. Lemma 3.9 shows that the sign is $(-1)^q$, which is $-\varepsilon_r$ because $q+1$ is the number of times the digit 1 appears in the dyadic expansion of r. □

3.3. **Proof of Theorem 1.7.** The following lemma provides the crucial step in the proof of Theorem 1.7. It shows, for large enough n, that if the distance of σ_{n-1} to τ is less than $g_k(n)$ then either this distance is also less than $g_{k+1}(n)$ or σ_{n-1+2^k} has distance from τ which is less than $g_k(n+2^k)$.

Lemma 3.11. *Let $k \ge 0$ and assume that $n \ge 2^{k+1}(k+1)$ is such that*

$$\sigma_{n-1} \le \tau < \sigma_{n-1} + g_k(n). \tag{3.3}$$

Then for $0 \le r < 2^{k+1}$ one has $s_{n+r} = \varepsilon_r$ and one of the following inequalities holds:

$$\sigma_{n-1} \le \tau < \sigma_{n-1} + g_{k+1}(n), \tag{3.4}$$

$$\sigma_{n-1+2^k} - g_k(n+2^k) \le \tau < \sigma_{n-1+2^k}. \tag{3.5}$$

Similarly, if $n \ge 2^{k+1}(k+1)$ is such that

$$\sigma_{n-1} - g_k(n) \le \tau < \sigma_{n-1}, \tag{3.6}$$

then for $0 \le r < 2^{k+1}$ one has $s_{n+r} = -\varepsilon_r$ and one of the following inequalities holds:

$$\sigma_{n-1} - g_{k+1}(n) \le \tau < \sigma_{n-1},$$
$$\sigma_{n-1+2^k} \le \tau < \sigma_{n-1+2^k} + g_k(n+2^k).$$

Proof. We shall consider only the case where (3.3) holds, the other case being analogous. Assuming (3.3), by Lemma 3.9 and Corollary 3.10 we have

$$\operatorname{sgn}\left(\sigma_{n-1} - \tau + \sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+\ell}\right) = \operatorname{sgn}\left(\sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+\ell}\right) = -\varepsilon_r \tag{3.7}$$

for all $r \in (0, 2^k)$. Moreover, if $\sigma_{n-1} \neq \tau$, then the claim holds also for $r = 0$, since $\operatorname{sgn}(\sigma_{n-1} - \tau) = -1 = -\varepsilon_0$. The equalities in (3.7) and the definition of the greedy algorithm then imply that, in any case,

$$\sigma_{n-1+r} = \sigma_{n-1} + \sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+\ell} \quad \text{for all } r \in [0, 2^k]. \tag{3.8}$$

This relation with $r = 2^k$ and the equality $\sum_{0\le\ell<2^k} \frac{\varepsilon_\ell}{n+\ell} = g_k(n)$ in Lemma 3.2 show that

$$\sigma_{n-1+2^k} = \sigma_{n-1} + \sum_{0\le\ell<2^k} \frac{\varepsilon_\ell}{n+\ell} = \sigma_{n-1} + g_k(n),$$

which is part of what we have to prove in this case. Further, this equality and (3.3) imply that

$$\sigma_{n-1+2^k} - g_k(n) \le \tau < \sigma_{n-1+2^k}. \tag{3.9}$$

In particular, this implies that

$$\sigma_{n+2^k} = \sigma_{n-1+2^k} - \frac{1}{n+2^k} = \sigma_{n-1+2^k} - \frac{\varepsilon_0}{n+2^k}. \tag{3.10}$$

By (3.9) and appealing once again to Lemma 3.9 and Corollary 3.10, we have

$$\operatorname{sgn}\left(\sigma_{n-1+2^k} - \tau - \sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+2^k+\ell}\right) = -\operatorname{sgn}\left(\sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+2^k+\ell}\right) = \varepsilon_r$$

for all $r \in (0, 2^k)$. Thus, since we have also (3.10) we must have

$$\sigma_{n-1+2^k+r} = \sigma_{n-1+2^k} - \sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+2^k+\ell} \quad \text{for all } r \in [0, 2^k]. \tag{3.11}$$

By (3.8), (3.10) and (3.11) we have

$$\sigma_{n-1+r} = \sigma_{n-1} + \sum_{0\le\ell<r} \frac{\varepsilon_\ell}{n+\ell} \quad \text{for all } r \in [0, 2^{k+1}].$$

This proves that $s_{n+r} = \varepsilon_r$ for $r = 0, \ldots, 2^{k+1} - 1$. Moreover, the case $r = 2^{k+1}$ yields

$$\sigma_{n-1+2^{k+1}} = \sigma_{n-1} + \sum_{0\le\ell<2^{k+1}} \frac{\varepsilon_\ell}{n+\ell} = \sigma_{n-1} + g_{k+1}(n). \tag{3.12}$$

Then we have two possibilities: either $\sigma_{n-1+2^{k+1}} > \tau$ or $\sigma_{n-1+2^{k+1}} \le \tau$. In the first case, by (3.3) we have (3.4). In the second case, we observe that since $\tau < \sigma_{n-1+2^k}$ by (3.9), then comparing (3.8) with $r = 2^k$ and (3.12), we get

$$\sigma_{n-1+2^k} - g_k(n+2^k) = \sigma_{n-1+2^{k+1}} \le \tau < \sigma_{n-1+2^k},$$

so (3.5) holds. □

Corollary 3.12. *Let $k \ge 0$ and let $n \ge 2^{k+1}(k+1)$. Suppose that we have $(s_m)_{n\le m<n+2^k} = B_k$ or $(s_n)_{n\le m<n+2^k} = -B_k$ with B_k as in Theorem 1.7. Then there exists a sequence $(\delta_i)_{i\ge0}$ with $\delta_i \in \{\pm1\}$ for all i such that*

$$(s_m)_{m\ge n} = (\delta_0 \cdot B_k, \delta_1 \cdot B_k, \delta_2 \cdot B_k, \ldots).$$

Proof. Suppose $(s_m)_{n\le m<n+2^k} = B_k$, the other case is proved in a similar way. Notice that it suffices to show that $(s_m)_{n+2^k\le m<n+2^{k+1}} = \pm B_k$, since one can then iterate the same argument. Since $s_n = \varepsilon_0 = 1$, we have $\sigma_{n-1} \le \tau$. Moreover, since $s_{n+2^{k-1}} = \varepsilon_{2^{k-1}} = -1$, by (2.3) and the definition of g_{k-1} (which is positive by Lemma 3.5), we deduce that

$$\sigma_{n-1} \le \tau < \sigma_{n-1} + g_{k-1}(n) = \sigma_{n+2^{k-1}-1},$$

where the second inequality is strict by (2.2). Thus the hypothesis (3.3) in Lemma 3.11 is satisfied with $k-1$ in place of k, whence either (3.4) or (3.5) holds with k in place of $k+1$. In the first case, we have that (3.3) is satisfied; hence, applying Lemma 3.11 once again, we obtain $s_{n+r} = \varepsilon_r$ for $0 \le r < 2^{k+1}$, i.e., $(s_m)_{n\le m<n+2^{k+1}} = B_{k+1} = (B_k, -B_k)$, and in particular $(s_m)_{n+2^k\le m<n+2^{k+1}} = -B_k$. In the second case, we notice that (3.5), with $k-1$ in place of k, is actually hypothesis (3.6) with $k-1$ in place of k and $n+2^k$ in place of n. Thus, applying Lemma 3.11 we obtain $(s_m)_{n+2^k\le m<n+2^{k+1}} = B_k$, as desired. □

We are now in a position to prove Theorem 1.7. First, we define the exceptional sets X_k. Also we recall that $Y_k := \cup_{h\le k} X_h$ with $Y_0 := \emptyset$.

Definition 3.13. *For all $k \in \mathbb{N}$ the set X_k is defined as*

$$X_k := \Big\{\tau \in \mathbb{R} \colon \exists N \ge 0 \text{ and } 0 \le m < 2^k \text{ s. t. } s_n = f_k(n+m)\ \forall n \ge N\Big\}, \tag{3.13}$$

where we recall that f_k denotes the 2^k-periodic function such that $f_k(n) = \varepsilon_n$ for all $n \in [0, 2^k)$.

Proof of Theorem 1.7. First, we observe that the result is tautological for $\tau \in X_k$. Indeed, by the definition of X_k there exists N such that

$$(s_n)_{n\ge N} = (s_1, s_2, \dots, s_{N-1}, B_k, B_k, B_k, \dots) = (s_1 \cdot B_0, s_2 \cdot B_0, \dots, s_{N-1} \cdot B_0, B_k, B_k, B_k, \dots).$$

Thus, let us assume that $\tau \notin Y_\infty$.

By Proposition 2.2 $(s_n)_n$ has infinitely many changes of signs. Equivalently, it contains infinitely many blocks B_1. In particular, we can find N_1 such that $N_1 \ge 2^{1+1} \cdot (1+1) = 8$ and $(s_m)_{N_1\le m<N_1+2} = B_1$. Thus, by Corollary 3.12 there exists a sequence $(\delta_{i,1})_{i\ge 0}$ with values in $\{\pm 1\}$ such that

$$(s_n)_{n\ge N_1} = (\delta_{0,1} \cdot B_1, \delta_{1,1} \cdot B_1, \delta_{2,1} \cdot B_1, \dots). \tag{3.14}$$

The sequence $(\delta_{i,1})_{i\ge 0}$ contains infinitely many changes of signs. Indeed, it cannot be equal to 1 for all i large enough, since otherwise we would have $\tau \in X_1$, and for the same reason it cannot be equal to -1 for all n large, since

$$(-B_1, -B_1, -B_1, \dots) = (-B_0, B_1, B_1, B_1, \dots).$$

Thus, we can find i such that $\delta_i = -\delta_{i+1} = 1$ and such that the corresponding index n on the left hand side of (3.14) is $N_2 \ge 2^{2+1} \cdot (2+1) = 24$ (and $N_2 > N_1$). We then have $(s_m)_{N_2\le m<N_2+2^2} = (B_1, -B_1) = B_2$ and so by Corollary 3.12 there exists a sequence $(\delta_{i,2})_{i\ge 0}$ with values in $\{\pm 1\}$ such that

$$(s_n)_{n\ge N_2} = (\delta_{0,2} \cdot B_2, \delta_{1,2} \cdot B_2, \delta_{2,2} \cdot B_2, \dots).$$

As before, one sees that $(\delta_{i,2})_{i\ge 0}$ contains infinitely many changes of signs. We can then keep iterating this process, whence obtaining sequences $(k_i)_i$ and $(\kappa_i)_i$ with the desired properties. □

4. The exceptional sets X_k

In this section we study the exceptional sets X_k, defined in (3.13).

As observed in the introduction, the elements in X_k can be written as a rational number plus $U_{k,m}$ for some $m \in [0, 2^k)$, where the constant $U_{k,m}$ is as defined in (1.8). We also recall that $U_{k,m} = -U_{k,m+2^{k-1}}$ for all $m \in [0, 2^{k-1})$. The values of $U_{k,m}$ for $k = 1, 2$ have been given in (1.9). For $k = 3$ we have

$$U_{3,0} = \frac{4\sqrt{2}\log(2-\sqrt{2}) - \pi - (\sqrt{2}+1)\log 4}{8}, \qquad U_{3,1} = \frac{(1-2\sqrt{2})\pi + \log 4}{8},$$
$$U_{3,3} = \frac{4\sqrt{2}\log(2-\sqrt{2}) + \pi - (\sqrt{2}+1)\log 4}{8}, \qquad U_{3,2} = \frac{(2\sqrt{2}-1)\pi + \log 4}{8}.$$

We now show that $U_{k,m}$ can be written as a linear combination of logarithms. Baker's theorem will then yield the transcendence of $U_{k,m}$.

Proposition 4.1. *For every $k \geq 1$, $m \in \mathbb{Z}$, we have*

$$U_{k,m} = \sum_{\substack{a=1 \\ a\,\text{odd}}}^{2^k-1} c(a) e^{2\pi i a m/2^k} \log\left(1 - e^{2\pi i a/2^k}\right), \qquad c(a) := i^k e^{\pi i a/2^k} \prod_{j=1}^{k} \sin\left(\frac{\pi a}{2^j}\right).$$

Proof. We write $f_k(n)$ in terms of additive characters:

$$f_k(n) = \frac{1}{2^k} \sum_{a=0}^{2^k-1} \left(\sum_{\ell=0}^{2^k-1} \varepsilon_\ell \, e^{-2\pi i \ell a/2^k} \right) e^{2\pi i a n/2^k}. \tag{4.1}$$

The sum in brackets is equal to $P_k(e^{-2\pi i a/2^k})$, where $P_k(x) := \sum_{\ell=0}^{2^k-1} \varepsilon_\ell x^\ell$. The uniqueness of the base-2 expansion gives the factorization $P_k(x) = \prod_{j=0}^{k-1}(1 - x^{2^j})$, so that

$$\frac{1}{2^k} \sum_{\ell=0}^{2^k-1} \varepsilon_\ell \, e^{-2\pi i \ell a/2^k} = \frac{1}{2^k} \prod_{j=0}^{k-1} \left(1 - e^{-2\pi i a/2^{k-j}}\right) = i^k \left(\prod_{j=0}^{k-1} e^{-\pi i a 2^j/2^k} \right) \prod_{j=1}^{k} \sin\left(\frac{\pi a}{2^j}\right)$$
$$= i^k e^{-\pi i a(2^k-1)/2^k} \prod_{j=1}^{k} \sin\left(\frac{\pi a}{2^j}\right) = -c(a),$$

where in the last step we used that the product on the second line is 0 unless a is odd.

Inserting (4.1) in the definition (1.8) of $U_{k,m}$ and exchanging the order of summation of the two sums, a step which can be easily justified, we obtain

$$U_{k,m} = -\sum_{\substack{a=1 \\ a\,\text{odd}}}^{2^k-1} c(a) e^{2\pi i a m/2^k} \sum_{n=1}^{+\infty} \frac{e^{2\pi i a n/2^k}}{n} = \sum_{\substack{a=1 \\ a\,\text{odd}}}^{2^k-1} c(a) e^{2\pi i a m/2^k} \log(1 - e^{2\pi i a/2^k}),$$

as claimed. □

Proposition 4.2. *The number $U_{k,m}$ is transcendental, for every $k \geq 1$ and every $m \in [0, 2^k)$.*

Proof. The formula in Proposition 4.1 shows that $U_{k,m}$ is a non-zero linear combination with algebraic coefficients of logarithms of the numbers $1 - \zeta^a$, where $\zeta := \exp(2\pi i/2^k)$ for $a = 1, \ldots, 2^k - 1$ odd. For any odd a, let $w_a := \zeta^{(1-a)/2}(1-\zeta^a)/(1-\zeta)$, which is well defined because $(1-a)/2$ is an integer. Notice that w_a is a positive real number for $1 \leq a < 2^{k-1}$, and that $w_{2^k - a} = -w_a$; also, $w_1 = 1$. It follows that the sum of the $\log(1-\zeta^a)$ is also a linear combination with algebraic coefficients of numbers

$$\log(1-\zeta), \qquad i\pi, \qquad \big(\log(w_a)\big)_{a=3,\text{odd}}^{2^{k-1}}. \tag{4.2}$$

The formula also shows that at least one of the coefficients is not zero; for example, the coefficient of $\log(1-\zeta)$ is $\sum_{a=1}^{2^k-1} c(a) e^{2\pi i a m/2^k} = -\varepsilon_m$. By Baker's theorem on linear forms in logarithms, the transcendence of $U_{k,m}$ then follows if we can prove that the numbers (4.2) are

$\mathbb{Q}$-linearly independent. Thus suppose we have a linear combination with integer coefficients producing zero:

$$\alpha\log(1-\zeta)+\beta i\pi+\sum_{\substack{a=3\\ a\,\mathrm{odd}}}^{2^{k-1}-1}\gamma_a\log(w_a)=0. \tag{4.3}$$

We need to show that all the coefficients are zero. Exponentiating this identity yields

$$(1-\zeta)^{\alpha}(-1)^{\beta}\prod_{\substack{a=3\\ a\,\mathrm{odd}}}^{2^{k-1}-1}w_a^{\gamma_a}=1. \tag{4.4}$$

In the cyclotomic field $K:=\mathbb{Q}[\zeta]$ of 2^k roots of unity, the norm of $1-\zeta$ is equal to 2, whereas the norm of each of the w_a is equal to 1. Thus the previous identity implies that $\alpha=0$. Taking the imaginary part of (4.3) then gives $\beta=0$. As a consequence (4.4) becomes

$$\prod_{\substack{a=3\\ a\,\mathrm{odd}}}^{2^{k-1}-1}w_a^{\gamma_a}=1. \tag{4.5}$$

It is known that the numbers $(w_a)_{a=3,\,a\,\mathrm{odd}}^{2^{k-1}-1}$ are cyclotomic units generating a subgroup having finite index in the free part of the group of units U_K in O_K [34, Lemma 8.1]. By Dirichlet's theorem the dimension (as $\mathbb{Z}$-module) of U_K is $2^{k-2}-1$, and this is also the number of w_a appearing in (4.5). Hence they are multiplicatively independent and so all the γ_a have to be equal to zero, as desired. $\square$

We will now show that $U_{k,m}\in X_k$ for all k,m. First, we need the following lemma.

Lemma 4.3. *For $k,m\geq 0$ and $x>0$, let*

$$g_{k,m}(x):=\sum_{0\leq n<2^k}\frac{f_k(n+m)}{n+x}.$$

Then $f_k(m)g_{k,m}(x)\geq g_{k,0}(x)=g_k(x)$ for $x>0$. In particular, $f_k(m)g_{k,m}(x)>0$ for $x>0$.

Proof. For $k\leq 1$ the result is obvious by definition, so assume that $k\geq 2$. Let

$$\mathcal{V}_{k,m}(r):=\sum_{0\leq n\leq r}f_k(n+m)$$

and notice that since $\mathcal{V}_{k,m}(2^k-1)=0$, one has that $\mathcal{V}_{k,m}(r+2^k)=\mathcal{V}_{k,m}(r)$ for all $r\geq 0$. Moreover, one easily verifies that

$$\mathcal{V}_{k,0}(r)=\begin{cases}f_k(r), & r \text{ even};\\ 0, & r\text{ odd};\end{cases}$$

(cf. Proposition 6.1 below). Thus, for all odd m we have

$$\begin{aligned}f_k(m)\mathcal{V}_{k,m}(r)&=f_k(m)(\mathcal{V}_{k,0}(r+m)-\mathcal{V}_{k,0}(m-1))=f_k(m)\mathcal{V}_{k,0}(r+m)+1\\ &=\begin{cases}1, & r\text{ even};\\ f_k(m)f_k(r+m)+1, & r\text{ odd};\end{cases}\end{aligned}$$

where in the second step we used that $f_k(m-1)f_k(m)=-1$ for m odd. In particular if m is odd, then

$$f_k(m)\mathcal{V}_{k,m}(r)-\mathcal{V}_{k,0}(r)\geq 0. \tag{4.6}$$

To conclude, we observe that, if $m = 2^\nu m' < 2^k$ with m' odd (or if $m = m' = 0$, $\nu < k$), then

$$g_{k,m}(x) = \sum_{n=0}^{2^k-1} \frac{f_k(n+m)}{n+x} = \sum_{\ell=0}^{2^\nu-1}\sum_{r=0}^{2^{k-\nu}-1} \frac{f_k(\ell+2^\nu(r+m'))}{\ell+2^\nu r+x} = \sum_{r=0}^{2^{k-\nu}-1} f_{k-\nu}(r+m')g_\nu(2^\nu r+x)$$

since (2.6) implies that $f_k(\ell+2^\nu s) = \varepsilon(\ell)f_k(2^\nu s) = \varepsilon(\ell)f_{k-\nu}(s)$ for $0 \le \ell < 2^\nu < 2^k$. Thus, by Abel's summation formula,

$$\begin{aligned} f_k(m)g_{k,m}(x) - g_{k,0}(x) &= \sum_{0\le r<2^{k-\nu}} \big(f_{k-\nu}(m')f_{k-\nu}(r+m') - f_{k-\nu}(r)\big)g_\nu(2^\nu r+x) \\ &= -2^\nu \int_0^{2^{k-\nu}-1} \big(f_{k-\nu}(m')\mathcal{V}_{k-\nu,m'}(y) - \mathcal{V}_{k-\nu,0}(y)\big)g_\nu'(2^\nu y+x)\,\mathrm{d}y \end{aligned}$$

and the result follows by (4.6) and Lemma 3.5. □

Corollary 4.4. *For all $k \ge 1$ and $m \in [1, 2^k)$ the identity $U_{k,m} = \sum_{n=1}^\infty f_k(n+m)/n$ gives the greedy representation for $U_{k,m}$, i.e., $s_n(U_{k,m}) = f_k(n+m)$ for all n. In particular, $U_{k,m} \in X_k$.*

Proof. By the definition of $U_{k,m}$, for all $N \ge 0$ we have

$$U_{k,m} = \sum_{n=1}^{N} \frac{f_k(n+m)}{n} + \sum_{n=N+1}^{+\infty} \frac{f_k(n+m)}{n} = \sum_{n=1}^{N} \frac{f_k(n+m)}{n} + \sum_{s=0}^{+\infty} g_{k,m+N+1}(N+1+s2^k).$$

Lemma 4.3 then shows that

$$f_k(m+N+1)\Big(U_{k,m} - \sum_{1\le n\le N} \frac{f_k(n+m)}{n}\Big) > 0 \tag{4.7}$$

for every N. This fact gives the claim by induction. Indeed, this inequality with $N = 0$ shows that

$$f_k(m+1) > 0 \iff U_{k,m} > 0 \iff s_1(U_{k,m}) > 0,$$

proving that $f_k(m+1) = s_1(U_{k,m})$. Moreover, if the claim holds for all $n \le N$, then $\sum_{n=1}^N \frac{f_k(n+m)}{n} = \sigma_N(U_{k,m})$ and (4.7) gives

$$f_k(m+N+1)(U_{k,m} - \sigma_N(U_{k,m})) \ge 0,$$

and so $s_{N+1}(U_{k,m}) = f_k(N+1+m)$, as desired. □

We record some properties of the sets X_k in the following proposition.

Proposition 4.5. *For all $k \ge 1$ we have $X_k \cap \overline{\mathbb{Q}} = \emptyset$. Moreover, the sets $X_1, X_2, \dots$ are non-empty, countable and pairwise disjoint.*

Proof. Since $U_{k,0} \in X_k \subseteq \mathbb{Q} + \{U_{k,m} \colon 0 \le m < 2^k\}$, the set X_k is non-empty, countable and, thanks to Proposition 4.2, we have $X_k \cap \overline{\mathbb{Q}} = \emptyset$. Now we show that $X_k \cap X_h = \emptyset$ for all $h \ne k$. Assume that $\tau \in X_k \cap X_h$ with $h \le k$. Thus, by definition, there exists m_1, m_2 such that $s_n = f_k(n+m_1) = f_h(n+m_2)$ for all sufficiently large n. In particular, $f_k(n)$ is periodic modulo 2^h. Since $f_k(0) = 1 = -f_k(2^{k-1})$, one has $h \ge k$ and so $h = k$. □

4.1. **Verifying whether $\tau \in X_k$ in a finite number of steps.** For each $k \ge 1$, the set X_k is defined as the set of all $\tau \in \mathbb{R}$ whose greedy sequence is eventually periodic with repeating block B_k. Since there is no effective bound on when the sequence starts being periodic, one might then expect that there is no algorithm which verifies whether a given number τ is in B_k. Surprisingly, however, such an algorithm can be constructed. We assume that $\tau \notin \mathbb{Q}$, since otherwise we already know $\tau \notin X_k$ for all k. Also, we assume that we know τ to arbitrary precision and that we can tell whether two real numbers coincide (actually, this is not needed if we already know that $\tau = r + U_{k,m}$ for some k, m and some $r \in \mathbb{Q}$).

The algorithm is rather simple and proceeds as follows.

(1) **Determine whether $\tau \in X_1$.** In order to determine if $\tau \in X_1$, one starts by finding the first $N_1 \geq 24$ such that $(s_m)_{N_1 \leq m < N_1 + 2^1} = B_1$. By Proposition 2.2 we know that such N_1 exists and by the upper bound for the harmonic sum we also know that $N_1 \leq \max(24, e^{|\tau|} + 2)$. Then we claim that one has $\tau \in X_1$ (and thus $\tau \notin X_k$ for all $k > 1$) if and only if

$$\tau = U_{1,m_1} + \sum_{1 \leq n < N_1} \frac{s_n - f_1(n + m_1)}{n}$$

where $0 \leq m_1 < 2^1$, $m_1 \equiv -N_1 \pmod 2$. In particular, the algorithm stops if this equality is satisfied and otherwise it moves to the next step. To prove this equivalence, we first observe that by Corollary 3.12 we have

$$(s_n)_{n \geq N_1} = (B_1, \delta_1 \cdot B_1, \delta_2 \cdot B_1, \dots)$$

with $\delta_i \in \{\pm 1\}$. Also, we have $\tau \in X_1$ if and only if $\delta_i = 1$ for all i. Indeed, if $\delta_i = 1$ for all i then by definition $\tau \in X_1$. Conversely, if $\delta_i = -1$ for some i then $(s_n)_{n \geq N_1}$ contains the block $(B_1, -B_1)$ and so, by Corollary 3.12,

$$(s_n)_{n \geq N'} = (B_2, \delta'_1 \cdot B_2, \delta'_2 \cdot B_2, \dots).$$

for some $N' \geq N_1$. In particular, the tail of $(s_n)_n$ cannot be periodic with repeating block B_1, i.e., $\tau \notin X_1$. Finally, one easily sees that $\delta_i = 1$ for all i, i.e.,

$$(s_n)_{n \geq N_1} = (B_1, B_1, B_1, \dots),$$

if and only if

$$\tau = \sum_{1 \leq n < N_1} \frac{s_n}{n} + \sum_{n \geq N_1} \frac{f_1(n + m_1)}{n} = \sum_{1 \leq n < N_1} \frac{s_n - f_1(n + m_1)}{n} + U_{1,m_1}.$$

One then proceeds inductively. Assuming $\tau \notin X_{k-1}$ with $k \geq 2$, to verify whether $\tau \in X_k$ one proceeds as follows.

(2) **Determine the first $N_k \geq 2^{k+2}(k+2)$ such that** $(s_m)_{N_k \leq m < N_k + 2^k} = B_k$. Since $\tau \notin X_{k-1}$ we know that such an N_k exists. We can also provide an explicit bound for it in terms of τ and $(s_n)_{n < N_{k-1}}$; we shall give the details at the end of the algorithm.

(3) **Determine whether $\tau \in X_k$.** In the same way as in step (1), one has that $\tau \in X_k$ if and only if

$$\tau = U_{k,m_k} + \sum_{1 \leq n < N_k} \frac{s_n - f_k(n + m_k)}{n}$$

where $0 \leq m_k < 2^k$, $m_k \equiv -N_k \pmod{2^k}$.

As anticipated, given $\tau \notin X_k$ and the first $N_{k-1} \geq 2^{k+1}(k+1)$ such that $(s_m)_{N_2 \leq m < N_2 + 2^{k-1}} = B_{k-1}$, one can give an upper bound on N_k. Indeed, since $\tau \notin X_{k-1}$ we know that

$$G_{\tau, N_{k-1}} := \tau - U_{k-1, m_{k-1}} + \sum_{n < N_{k-1}} \frac{f_{k-1}(n + m_{k-1}) - s_n}{n} \neq 0.$$

Then we have $N_k \leq \max\left(N_{k-1}, 2^{k+2}(k+2), 2^k + 4/|G_{\tau, N_{k-1}}|\right)$. To see this, we observe that, by definition, for any $M \geq N_{k-1}$ we have

$$G_{\tau, N_{k-1}} = \sum_{N_{k-1} \leq n < M} \frac{s_n - f_{k-1}(n + m_1)}{n} + \sum_{n \geq M} \frac{s_n}{n} - \sum_{n \geq M} \frac{f_{k-1}(n + m_{k-1})}{n}.$$

In particular,

$$\sum_{N_{k-1} \leq n < M} \frac{|s_n - f_{k-1}(n + m_{k-1})|}{n} > |G_{\tau, N_{k-1}}| - \left| \sum_{n \geq M} \frac{s_n}{n} \right| - \left| \sum_{n \geq M} \frac{f_{k-1}(n + m_{k-1})}{n} \right|. \tag{4.8}$$

By Proposition 2.2 and Corollary 4.4 both sums on the right are bounded by $2/M$. Thus, (4.8) implies

$$\sum_{N_{k-1}\le n<M} \frac{|s_n - f_{k-1}(n+m_{k-1})|}{n} > |G_{\tau,N_{k-1}}| - \frac{4}{M}.$$

Taking M to be the largest integer such that $M \equiv N_{k-1} \pmod{2^{k-1}}$ and $M < 2^k + 4/|G_{\tau,N_{k-1}}|$, we have that the right hand side is greater than zero. Thus, with this choice we obtain

$$\sum_{N_{k-1}\le n<M} \frac{|s_n - f_{k-1}(n+m_{k-1})|}{n} > 0$$

and so $s_n \neq f_{k-1}(n+m_{k-1})$ for some $n \in [N_{k-1}, M)$. Since $(s_n)_{n\ge N_{k-1}}$ can be written in terms of blocks $\pm B_{k-1}$ it follows that $(s_n)_{N_{k-1}\le n<M}$ has to contain a block $(B_{k-1}, -B_{k-1}) = B_k$.

Remark 4.6. To give some examples, using the above algorithm we verified that $U_{2,0}+r \in X_2$ if $r = 1, 2, 3$ and $U_{2,0} + r \notin X_2$ if $r = 4, \dots, 10$.

5. Proof of Theorem 1.5

Definition 5.1. *For $\tau \notin Y_\infty$ and $h \ge 1$, let n_h be the minimum integer such that $|\tau - \sigma_{n_h-1}(\tau)| < g_{h-1}(n_h)$ and $n_h \ge 2^h h$. Notice that by Corollary 1.8 and* (2.3) *we know that such an integer exists.*

The following lemma gives some information on the sequence $(n_h)_h$ and in particular it put a rather sharp limit for the possible sign patterns in (s_n) for $n_h \le n < n_{h+1}$.

Lemma 5.2. *Assume that $\tau \notin (Y_\infty \cup \mathbb{Q})$. The sequence $(n_h)_h$ is non-decreasing and such that $n_{h+1} \equiv n_h \pmod{2^h}$ for all $h \ge 1$; also, $n_1 \le 4e^{|\tau|}$. Moreover, if $n_h < n_{h+1}$ we have*

$$(s_n)_{n_h\le n<n_{h+1}} = (\eta_{0,h}\cdot B_h, \eta_{1,h}\cdot B_h, \dots, \eta_{r_h,h}\cdot B_h),$$

where $r_h := (n_{h+1} - n_h)/2^h - 1$ and $\eta_{i,h} \in \{\pm1\}$ is such that $\eta_{i,h} = s_{n_{h+1}}$ for all $i \in [0, r_h]$ if $n_h \ge 2^{h+1}(h+1)$ and for all $i \in [h+2, r_h]$ in any case.

Proof. The fact that n_h is non-decreasing follows immediately from Corollary 3.6, whereas the inequality $n_1 \le \max(e^{|\tau|} + 1, 4) \le 4e^{|\tau|}$ follows from bounding the harmonic sum.

By Lemma 3.11 and Corollary 3.12 we have

$$(s_n)_{n\ge n_h} = (\eta_{0,h}\cdot B_h, \eta_{1,h}\cdot B_h, \dots),$$

for some $(\eta_{i,h})_{i\ge0}$ with $\eta_{i,h} \in \{\pm1\}$. By Lemma 3.11 we have $(s_n)_{n_{h+1}\le n<n_{h+1}+2^{h+1}} = B_{h+1}$ and so it must be $n_{h+1} = n_h + 2^h r_h$ for some $r_h \ge 0$.

We claim that if $\eta_{j,h} \neq \eta_{j+1,h}$ for some $j \in [0, r_h]$ (with $\eta_{r_h+1,h} := \eta_{0,h+1}$), then $n_h + 2^h j < 2^{h+1}(h+1)$. Indeed, if $\eta_{j,h} \neq \eta_{j+1,h}$, then Equation (2.3) implies that $|\sigma_{n'-1} - \tau| < g_h(n')$ with $n' := n_h + 2^h j$ (the inequality is strict since $\tau \notin \mathbb{Q}$). Since $n' < n_{h+1}$ it then follows that it must be $n' = n_h + 2^h j < 2^{h+1}(h+1)$, as claimed. This implies in particular that if $n_h \ge 2^{h+1}(h+1)$ then $\eta_{i,h} = \eta_{i+1,h}$ for $0 \le i \le r_h$ or, equivalently, $\eta_{i,h} = s_{n_{h+1}}$ for $0 \le i \le r_h$ since $\eta_{r_h+1,h} = s_{n_{h+1}}$. Similarly, if $h+2 \le i \le r_h$ then $n_h + 2^h i \ge 2^h h + (h+2)2^h = 2^{h+1}(h+1)$ and so one concludes as before. □

Thanks to the above lemma we have a good control on the sequence $(n_h)_h$. We now need two combinatorial lemmas. The first one is a well known upper bound for the number of lattice point in a simplex [10]. We give a simple proof for completeness.

Lemma 5.3. *Let $k, m \ge 1$ and $b_1, \dots, b_k \in \mathbb{N}$. Then*

$$S := |\{(a_1, \dots, a_k) \in \mathbb{Z}^k_{\ge0} \colon b_1a_1 + b_2a_2 + \cdots + b_ka_k = m\}| \le \frac{(m + b_2 + \cdots + b_k)^{k-1}}{(b_2\cdots b_k)(k-1)!}.$$

Proof. Clearly,

$$\begin{aligned} S &\le |\{(a_2,\dots,a_k)\in\mathbb{Z}_{\ge 0}^{k-1} : b_2a_2+\cdots+b_ka_k\le m\}| \\ &\le \int_0^\infty\cdots\int_0^\infty \chi_{[0,1]}\Big(\frac{b_2x_2+\cdots+b_kx_k}{m+b_2+\cdots+b_k}\Big)\,\mathrm{d}x_2\cdots\,\mathrm{d}x_k. \end{aligned}$$

By a change of variables this is

$$\frac{(m+b_2+\cdots+b_k)^{k-1}}{b_2\cdots b_k}\int_0^\infty\cdots\int_0^\infty \chi_{[0,1]}(x_2+\cdots+x_k)\,\mathrm{d}x_2\cdots\,\mathrm{d}x_k = \frac{(m+b_2+\cdots+b_k)^{k-1}}{(b_2\cdots b_k)(k-1)!},$$

as desired. □

Lemma 5.4. *For $k\ge 3$, $\ell\ge 0$ we have*

$$Z(k,\ell) := \sum_{\substack{a_1,\dots,a_{k-1}\ge 0\\ a_1+\cdots+2^{k-2}a_{k-1}=\ell}}\ \prod_{2\le h\le k} 2^{h\chi_{[0,2^h h)}(a_1+2a_2+\cdots+2^{h-2}a_{h-1})} \le 2^{-\frac{k^2}{2}+5k}\frac{\ell^{k-2}}{(k-2)!}+2^{\frac{k^2}{2}+4k}k!.$$

Proof. First, we observe that $Z(k,\ell)\le 4Z'(k,\ell/2)$, where

$$Z'(k,\ell) = \sum_{\substack{a_2,\dots,a_{k-1}\ge 0\\ a_2+\cdots+2^{k-3}a_{k-1}\le\ell}}\ \prod_{3\le h\le k} 2^{h\chi_{[0,2^{h-1}h)}(a_2+\cdots+2^{h-3}a_{h-1})}.$$

We shall prove that for all $k\ge 3,\ell\ge 0$ one has

$$Z'(k,\ell)\le 2^{-\frac{k^2}{2}+5k-2}\frac{\ell^{k-2}}{(k-2)!}+2^{\frac{k^2}{2}+4k-2}k!, \tag{5.1}$$

from which the claimed inequality for Z follows immediately. We prove this by induction over k. The result is obvious if $k=3$. Now assume that the claimed inequality holds for $k-1\ge 3$. We split Z' into $Z'=Z'_<+Z'_\ge$, where $Z'_<$ is the contribution of the terms with $a_2+\cdots+2^{k-3}a_{k-1}<2^{k-1}k$. We have

$$\begin{aligned} Z'_<(k,\ell) &\le 2^k \sum_{\substack{a_2,\dots,a_{k-1}\ge 0,\ a_{k-1}<4k\\ a_2+\cdots+2^{k-4}a_{k-2}<2^{k-3}(4k-a_{k-1})}}\ \prod_{3\le h\le k-1} 2^{h\chi_{[0,2^{h-1}h)}(a_2+\cdots+2^{h-3}a_{h-1})} \\ &= 2^k\sum_{0\le a<4k} Z'(k-1,2^{k-3}(4k-a)). \end{aligned}$$

Thus, by the inductive hypothesis,

$$\begin{aligned} Z'_<(k,\ell) &\le \sum_{0\le a<4k} 2^{\frac{k^2+2k+3}{2}}\frac{(4k-a)^{k-3}}{(k-3)!}+2^{\frac{k^2+6k-11}{2}}2^k4k(k-1)! \\ &\le \frac{2^{\frac{k^2+2k+3}{2}}}{(k-3)!}\Big[\int_0^{4k}(4k-x)^{k-3}\,\mathrm{d}x+(4k)^{k-3}\Big]+2^{\frac{k^2+8k-7}{2}}k! \\ &= 2^{\frac{k^2+6k-9}{2}}\frac{k^{k-3}}{(k-3)!}\Big[\frac{4k}{k-2}+1\Big]+2^{\frac{k^2+8k-7}{2}}k! \\ &\le 2^{\frac{k^2+9k-9}{2}}+2^{\frac{k^2+8k-7}{2}}k!, \end{aligned} \tag{5.2}$$

since $\frac{k^{k-3}}{(k-3)!}\big[\frac{4k}{k-2}+1\big] \leq 2^{3k/2}$ for all $k \geq 4$. In order to bound $Z'_{\geq}$, we first observe we can assume that $\ell \geq 2^{k-1}k$, since otherwise $Z'_{\geq}$ is just the empty sum. Now

$$\begin{aligned}
Z'_{\geq}(k,\ell) &\leq \sum_{0\leq a_{k-1}\leq \ell/2^{k-3}} \sum_{\substack{a_2,\dots,a_{k-2}\geq 0\\ a_2+\cdots+2^{k-4}a_{k-2}\leq \ell-2^{k-3}a_{k-1}}} \prod_{3\leq h\leq k-1} 2^{h\chi_{[0,2^{h-1}h)}(a_2+\cdots+2^{h-3}a_{h-1})}\\
&\leq \sum_{0\leq a_{k-1}\leq \ell/2^{k-3}} Z'(k-1,\ell-2^{k-3}a_{k-1}).
\end{aligned}$$

Thus, by the inductive hypothesis we have

$$\begin{aligned}
Z'_{\geq}(k,\ell) &\leq \sum_{0\leq a\leq \ell/2^{k-3}} \frac{(\ell-2^{k-3}a)^{k-3}}{(k-3)!}2^{-\frac{k^2-12k+15}{2}} + 2^{\frac{k^2+6k-11}{2}}(\ell/2^{k-3}+1)(k-1)!\\
&\leq \frac{2^{-\frac{k^2-12k+15}{2}}}{(k-3)!}\bigg(\int_0^{\ell/2^{k-3}}(\ell-2^{k-3}x)^{k-3}\,\mathrm{d}x+\ell^{k-3}\bigg)+\big(2^{\frac{k^2+4k-5}{2}}\ell+2^{\frac{k^2+6k-11}{2}}\big)(k-1)!\\
&= 2^{-\frac{k^2-10k+15}{2}}\frac{\ell^{k-2}}{(k-2)!}\Big[8+\frac{2^k(k-2)}{\ell}\Big]+\big(2^{\frac{k^2+4k-5}{2}}\ell+2^{\frac{k^2+6k-11}{2}}\big)(k-1)!\\
&\leq 10\cdot 2^{-\frac{k^2-10k+15}{2}}\frac{\ell^{k-2}}{(k-2)!}+\big(2^{\frac{k^2+4k-5}{2}}\ell+2^{\frac{k^2+6k-11}{2}}\big)(k-1)!, \qquad (5.3)
\end{aligned}$$

since $\ell \geq 2^{k-1}k$. Now if $\ell \geq 2^{2k-1}k$ then

$$2^{\frac{k^2+4k-5}{2}}\ell(k-1)! \leq 2^{\frac{k^2+4k-5}{2}}\ell^{k-2}\frac{2^{-2k(k-3)}k!}{k^{k-3}} \leq 6\ell^{k-2}2^{-\frac{k^2-10k+5}{2}}2^{-k^2+3k} \leq 2^{-\frac{k^2-10k+5}{2}}\frac{\ell^{k-2}}{(k-2)!}$$

since $k! \leq 6k^{k-3}$ and $6(k-2)! \leq 2^{k^2-3k}$ for $k \geq 4$. Thus, in any case

$$2^{\frac{k^2+4k-5}{2}}\ell(k-1)! \leq 2^{-\frac{k^2-10k+5}{2}}\frac{\ell^{k-2}}{(k-2)!}+2^{\frac{k^2+8k-7}{2}}k!$$

and so (5.1) follows from (5.2) and (5.3) since $2^{\frac{k^2+9k-9}{2}}+2^{\frac{k^2+6k-11}{2}}(k-1)! \leq 2^{\frac{k^2+8k-8}{2}}k!$. $\square$

The following two propositions give the crucial steps in proving Theorem 1.5. The first one essentially shows that, for almost all τ, the size of n_k is about $2^k k$ for infinitely many k. The second one shows that for almost all τ there are no $n \asymp 2^k$ with $|\tau-\sigma_{n-1}(\tau)|$ much smaller than $g_k(n)$.

Proposition 5.5. *Let $f:\mathbb{R}_{>0}\to\mathbb{R}_{>0}$ be a function such that $f(x)\to\infty$ as $x\to\infty$ and let*

$$\mathcal{X}_f := \left\{\tau\in\mathbb{R}\colon \begin{array}{l}\text{there are at most finitely many } k \text{ such that}\\ |\tau-\sigma_{n-1}(\tau)|<g_k(n) \text{ for some } n\in[2^k k, f(k)2^k k]\end{array}\right\}.$$

Then $\mathrm{meas}(\mathcal{X}_f)=0$.

Proof. Let $\mathcal{X}'_f := \mathcal{X}_f\cap\mathbb{R}'$ where $\mathbb{R}' := \mathbb{R}\setminus(Y_\infty\cup\mathbb{Q})$. By Proposition 4.5 we have that $\mathrm{meas}(\mathcal{X}_f)=\mathrm{meas}(\mathcal{X}'_f)$. Then we observe that

$$\begin{aligned}
\mathcal{X}'_f &\subseteq \bigcup_{m\in\mathbb{N}}\Big\{\tau\in\mathbb{R}'\colon |\tau-\sigma_{n-1}(\tau)|>g_k(n) \text{ for all } k\in\mathbb{N} \text{ and for all } n\in[2^k k, f(k)2^k k/m]\Big\}\\
&= \bigcup_{m\in\mathbb{N}}\bigcup_{q\in\mathbb{N}}\bigcap_{k\in\mathbb{N}}\mathcal{X}_{m,q,k},
\end{aligned}$$

where

$$\mathcal{X}_{m,q,k} := \left\{\tau\in\mathbb{R}'\cap[-q,q]\colon \begin{array}{l}|\tau-\sigma_{n-1}(\tau)|>g_h(n) \text{ for all } h\in[1,k]\\ \text{and for all } n\in[2^h h, f(h)2^h h/m]\end{array}\right\}.$$

In particular, in order to prove that $\mathrm{meas}(\mathcal{X}_f)=0$ it is sufficient to show that $\mathrm{meas}(\mathcal{X}_{m,q,k})\to 0$ as $k\to\infty$.

Let $\tau \in \mathcal{X}_{m,q,k}$ and let $n_1, n_2, \dots$ as in Lemma 5.2; in particular, $n_1 \leq 4e^q$. Also, let C be such that $f(x) \geq 4m$ for $x \geq C$ so that if $h \geq C$ then, by the definition of $\mathcal{X}_{m,q,k}$, $n_h \geq 2^{h+1}(h+1)$.

We split $\mathcal{X}_{m,q,k}$ depending on the value of n_k:

$$\mathcal{X}_{m,q,k} = \bigcup_{\ell \geq f(k)2^k k/m} V_{\ell,m,q,k}, \qquad V_{\ell,m,q,k} := \{\tau \in \mathcal{X}_{m,q,k} \colon n_k = \ell\}.$$

If $\tau \in V_{\ell,m,q,k}$, then, by definition, $|\tau - \sigma_{\ell-1}(\tau)| < g_{k-1}(\ell)$ and so τ has distance less than $g_{k-1}(\ell)$ from one of the elements of the set

$$R_{\ell,m,q,k} := \{\sigma_{\ell-1}(\tau) \colon \tau \in V_{\ell,m,q,k}\}.$$

This set has cardinality bounded by the number of possible choices of signs $(s_n)_{n<\ell}$. By Lemma 5.2, if $h \geq C$ then the signs in $(s_n)_{n_h \leq n < n_{h+1}}$ are completely determined (depending on the sign of $s_{n_{h+1}}$ and thus eventually on $s_{n_k} = s_\ell$), whereas for each $h < C$ the first $h+2 \leq C+1$ signs in $(s_n)_{n_h \leq n < n_{h+1}}$ are free and the other ones are determined. Also, there are at most $2^{n_1} \leq 2^{4e^q}$ possibilities for $(s_n)_{n<n_1}$ and 2 for s_ℓ. Finally, we recall that $n_1 \leq n_2 \leq \cdots \leq n_k = \ell$ and that $n_{h+1} \equiv n_h \pmod{2^h}$ for all $h \geq 1$. We then obtain that $|R_{\ell,m,q,k}|$ is bounded by $2^{C(C+1)+4e^q+1}$ times

$$\begin{aligned}
&|\{(m_1, \dots, m_{k-1}) \colon m_1 \leq \cdots \leq m_k := \ell,\ m_1 \leq 4e^q,\ m_{h+1} \equiv m_h \pmod{2^h} \text{ for } 1 \leq h < k\}| \\
&\quad = |\{(a_0, \dots, a_{k-1}) \in \mathbb{Z}_{\geq 0} \colon a_0 + 2a_1 + 2^2 a_2 + \cdots + 2^{k-1}a_{k-1} = \ell,\ a_0 \leq 4e^q\}| \\
&\quad \leq 4e^q \frac{(\ell + 2^2 + \cdots + 2^{k-1})^{k-2}}{(2^2 \cdots 2^{k-1})(k-2)!} \leq 4e^q 2 \frac{(\ell + 2^k)^{k-2}}{2^{\binom{k}{2}}(k-2)!} \ll e^q \frac{\ell^{k-2}}{2^{\binom{k}{2}}(k-2)!},
\end{aligned}$$

by Lemma 5.3, where in the last inequality we used that $\ell = n_k \geq 2^k k$. It follows that

$$\begin{aligned}
\operatorname{meas}(\mathcal{X}_{m,q,k}) &\leq 2 \sum_{\ell \geq f(k)2^k k/m} |R_{\ell,m,q,k}| g_{k-1}(\ell) \ll_{q,m} \sum_{\ell \geq f(k)2^k k/m} \frac{\ell^{k-2}}{2^{\binom{k}{2}}(k-2)!} \cdot \frac{2^{\binom{k-1}{2}}(k-1)!}{\ell^k} \\
&= \sum_{\ell \geq f(k)2^k k/m} \frac{k-1}{2^{k-1}\ell^2} \ll_{q,m} \frac{1}{2^{2k} f(k)},
\end{aligned}$$

by Lemma 3.5 (and for $k \geq C$). Thus, $\operatorname{meas}(\mathcal{X}_{m,q,k}) \to 0$ as $k \to \infty$, as desired. □

Proposition 5.6. *Let $f \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ be a function such that $f(x) \gg 2^{5x} x^x$ for x large enough and let*

$$\mathcal{Y}_f := \left\{ \tau \in \mathbb{R} \colon \begin{array}{l} \text{there exist infinitely many } k \text{ such that} \\ |\tau - \sigma_{n-1}(\tau)| < g_k(n)/f(k) \text{ for some } n \in [2^{k-1}(k-1), 2^k k) \end{array} \right\}.$$

Then $\operatorname{meas}(\mathcal{Y}_f) = 0$.

Proof. Let $\mathcal{Y}'_f := \mathcal{Y}_f \cap \mathbb{R}'$, with $\mathbb{R}' := \mathbb{R} \setminus (Y_\infty \cup \mathbb{Q})$, so that $\operatorname{meas}(\mathcal{Y}'_f) = \operatorname{meas}(\mathcal{Y}_f)$. We have

$$\mathcal{Y}'_f = \bigcup_{q \in \mathbb{N}} \bigcap_{K \in \mathbb{N}} \bigcup_{k \geq K} \bigcup_{2^{k-1}(k-1) \leq \ell < 2^k k} \mathcal{Y}_{q,k,\ell},$$

where

$$\mathcal{Y}_{q,k,\ell} := \{\tau \in \mathbb{R}' \cap [-q,q] \colon |\tau - \sigma_{\ell-1}(\tau)| < g_k(\ell)/f(k)\}.$$

It suffices to show that

$$\operatorname{meas}\Bigg(\bigcup_{k \geq K} \bigcup_{2^{k-1}(k-1) \leq \ell < 2^k k} \mathcal{Y}_{q,k,\ell} \Bigg) \to 0$$

as $K \to \infty$. Also, we can assume $f(x) \geq 1$ for $x \geq K$.

Let $\tau \in \mathcal{Y}_{q,k,\ell}$ and let $n_1, n_2, \dots$ be as in Lemma 5.2. By Corollary 3.6 we have $n_{k-1} \leq \ell < n_k$. Notice also that Lemma 5.2 implies that $\ell \equiv n_{k-1} \pmod{2^{k-1}}$.

The number τ belongs to $\mathcal{Y}_{q,k,\ell}$, so it has distance less than $\frac{g_k(\ell)}{f(k)}$ from one of the elements of the set

$$R_{q,k,\ell} := \{\sigma_{\ell-1}(\tau)\colon \tau \in \mathcal{Y}_{q,k,\ell}\}.$$

This set has cardinality bounded by the number of possible choices of signs $(s_n)_{n<\ell}$. By Lemma 5.2, for each $h \in [1, k-1]$ we have only one choice for the signs in $(s_n)_{n_h \le n < n_{h+1}}$ if $n_h \ge 2^{h+1}(h+1)$ and at most 2^{h+2} in any case. Also, we have at most 2^{4e^q} for the signs in $(s_n)_{n<n_1}$. Thus, writing $\ell = n_{k-1} + 2^{k-1}a_{k-1}$, $n_h = n_{h-1} + 2^{h-1}a_{h-1}$ for $2 \le h \le k-1$ and $n_1 = a_0$ for some $a_0, \dots a_k \in \mathbb{Z}_{\ge 0}$, we see that

$$\begin{aligned}|R_{q,k,\ell}| &\ll_q \sum_{\substack{a_0,a_1,\dots,a_{k-1}\ge 0\\ a_0+2a_1+2^2a_2+\cdots+2^{k-1}a_{k-1}=\ell}} \prod_{2\le h<k} 2^{(h+2)\chi_{[0,2^{h+1}(h+1))}(a_0+2a_1+2^2a_2+\cdots+2^{h-1}a_{h-1})}\\ &\ll_q 2^k \sum_{\substack{b_1,\dots,b_k\ge 0\\ b_1+2b_2+\cdots+2^{k-1}b_k=\ell}} \prod_{3\le h\le k} 2^{h\chi_{[0,2^h h)}(b_1+2b_2+\cdots+2^{h-2}b_{h-1})}.\end{aligned}$$

Applying Lemma 5.4 we then obtain

$$|R_{q,k,\ell}| \ll_q 2^{-\frac{k^2}{2}+5k}\frac{\ell^{k-1}}{(k-1)!} + 2^{\frac{k^2}{2}+5k}(k+1)! \quad \le 2^{\frac{k^2}{2}+4k}\frac{k^k}{k!} + 2^{\frac{k^2}{2}+6k}(k+1)! \ll 2^{\frac{k^2}{2}+5k}k^{k+1}$$

for $\ell \le 2^k k$ and $k \ge 3$ since $k! \ll k^k 2^{-k}$. In particular, for $K \ge 3$ we have

$$\begin{aligned}\operatorname{meas}\Bigg(\bigcup_{k\ge K}\ \bigcup_{2^{k-1}(k-1)\le \ell < 2^k k} \mathcal{Y}_{q,k,\ell}\Bigg) &\le \sum_{k\ge K}\ \sum_{2^{k-1}(k-1)\le\ell<k2^k} |R_{q,k,\ell}|\frac{g_k(\ell)}{f(k)}\\ &\ll_q \sum_{k\ge K}\ \sum_{2^{k-1}(k-1)\le\ell<k2^k} 2^{\frac{k^2}{2}+5k}k^{k+1}\frac{2^{\binom{k}{2}}k!}{f(k)\ell^{k+1}}\\ &\ll \sum_{k\ge K} 2^{11k/2}\frac{k!\,k^k}{(k-1)^k f(k)} \ll \sum_{k\ge K}\frac{2^{9k/2}k^k}{f(k)},\end{aligned}$$

by Lemma 3.5. The sum goes to zero if $f(x) \gg 2^{5x}x^x$ and so the proof is complete. ∎

The following lemmas put a lower and an upper limit to how small $|\tau - \sigma_{m-1}(\tau)|$ can be when τ is in $\mathbb{R}\setminus\mathcal{Y}_f$ and $\mathbb{R}\setminus\mathcal{X}_f$ respectively.

Lemma 5.7. *Let $f\colon \mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$ be an increasing function with $\log(f(x)) \le x\log x + O(x)$ as $x \to \infty$. Then for all $\tau \notin \mathcal{Y}_f$ we have*

$$|\tau - \sigma_{m-1}(\tau)| \ge e^{-\frac{(\log m)^2}{\log 4}+O_\tau(\log m)}$$

as $m \to \infty$.

Proof. If $\tau \notin \mathcal{Y}_f$ then there exists a sufficiently small $\delta > 0$ such that $|\tau - \sigma_{n-1}(\tau)| > \delta\frac{g_k(n)}{f(k)}$ for all $k \ge 1$ and all $n \in [2^{k-1}(k-1), 2^k)$. Now take any $m \ge 2$ and let $k \in \mathbb{N}$ be such that $m \in [2^k k, 2^{k+1}(k+1))$. Clearly such a k exists and satisfies $k+\log_2 k \le \log_2 m \le (k+1)+\log_2(k+1)$; in particular it follows that

$$k = \log_2 m - \log_2\log_2 m + O(1) = \frac{\log m - \log\log m}{\log 2} + O(1)$$

as $m \to \infty$. The result then follows, since by Lemma 3.5 we have

$$|\tau - \sigma_{m-1}(\tau)| > \delta\frac{g_k(m)}{f(k)} > \frac{\delta\, k!\, 2^{\binom{k}{2}}}{f(k)(2^{k+1}(k+1))^{k+1}} = \frac{e^{-\frac{\log 2}{2}k^2+O_\tau(k)}}{f(k)} > e^{-\frac{(\log m)^2}{\log 4}+O_\tau(\log m)}.$$

∎

Lemma 5.8. *Let $f\colon \mathbb{R}_{>0}\to\mathbb{R}_{>0}$ be an increasing function with $f(x)\geq 3$ for x large enough. Then for all $\tau\notin\mathcal{X}_f$ there exist arbitrary large n such that*

$$|\tau-\sigma_{m-1}(\tau)|\leq e^{-\frac{(\log m)^2-2\log m\log\log m}{\log 4}+O_\tau(\log m\log(f(\log n)))}$$

as $m\to\infty$.

Proof. Let $\tau\notin\mathcal{X}_f$. Then there exists arbitrarily large $k,n\in\mathbb{N}$ such that $|\tau-\sigma_{n-1}(\tau)|<g_k(n)$ and $2^k k\leq n\leq f(k)2^k k$. In particular, $k\leq\log n$ and we have

$$k=\frac{\log n-\log\log n}{\log 2}+O(\log(f(\log n))).$$

One then concludes as for Lemma 5.7. □

Proof of Theorem 1.5. Let $\mathcal{Y}:=\mathbb{R}\setminus(\mathcal{X}_f\cup\mathcal{Y}_g)$ with $f(x)=\log x$ and $g(x)=2^{5x}x^x$. By Propositions 5.5 and 5.6 we have $\operatorname{meas}(\mathcal{X}_f\cup\mathcal{Y}_g)=0$, whereas from Lemma 5.7 and Lemma 5.8 we deduce that, for all $\tau\in\mathcal{Y}$,

$$0\leq\liminf_{m\to\infty}\frac{\log|\tau-\sigma_{m-1}(\tau)|+\frac{1}{\log 4}(\log m)^2}{\log m\log\log m}\leq\frac{1}{\log 2},$$

which is Theorem 1.5 in a stronger form. □

We conclude the section with the following proposition, which implies, in particular, that the conclusion of Theorem 1.5 does not hold for all $\tau\in\mathbb{R}$.

Proposition 5.9. *Given $f\colon\mathbb{N}\to\mathbb{R}_{>0}$ we can construct $\tau_f\in\mathbb{R}$ such that $|\tau_f-\sigma_n(\tau_f)|\leq f(n)$ for infinitely many n. Moreover, the set of τ having this property is dense in $\mathbb{R}$.*

Proof. For simplicity we will only prove that there exists a τ_f with the required property. Along the same lines one can show that the set of τ with this property is dense in $\mathbb{R}$.

We first observe that we can assume $f(n+1)<\frac14 f(n)$ for all $n\geq 1$ and that $f(1)<1$. Next we observe that it suffices to construct a sequence of irrational numbers $(\tau_i)_i$ and an increasing sequence of integers $(m_i)_i$ such that, for all $i\geq 1$,

$$\sigma_n(\tau_{i+1})=\sigma_n(\tau_i)\ \forall n\leq m_i,\qquad |\sigma_{m_i}(\tau_i)-\tau_i|<\tfrac12 f(m_i),\qquad |\tau_{i+1}-\tau_i|<\tfrac14 f(m_i). \tag{5.4}$$

Indeed, this implies that for all $j>i$

$$|\tau_j-\tau_i|<\tfrac14(f(m_i)+\cdots+f(m_{j-1}))<\tfrac12 f(m_i)$$

and thus

$$|\tau_j-\sigma_{m_i}(\tau_i)|\leq|\tau_i-\sigma_{m_i}(\tau_i)|+|\tau_j-\tau_i|<f(m_i)$$

for all $i\leq j$. It follows that the limit $\tau_f:=\lim_{i\to\infty}\tau_i$ exists and satisfies $|\tau_f-\sigma_{m_i}(\tau_f)|\leq f(m_i)$ for all i, as desired.

We thus just have to see that such sequences exist. To construct the two sequences one proceed as follows. For τ_1 one takes $m_1=1$ and any irrational τ_1 in $(1-\frac12 f(1),1+\frac12 f(1))$. Clearly, one has $|\tau_1-\sigma_1(\tau_1)|<f(1)$. Now assume that we have two sequences $(\tau_r)_{r\leq i}$ and $(m_r)_{r\leq i}$ satisfying (5.4); we need to construct τ_{i+1} and m_{i+1}. For any $\tau\in\mathbb{R}$ and $m\in\mathbb{N}$ the set

$$I_m(\tau):=\{\alpha\in\mathbb{R}\colon\sigma_n(\alpha)=\sigma_n(\tau),\ \forall n\leq m\}$$

is an interval of non-zero measure containing τ. Since $\sigma_m(\tau_i)\to\tau_i$ as $m\to\infty$ we can find $q>m_i$ such that $\sigma_q(\tau_i)$ is in the interior of $I_{m_i}(\tau_i)$, $|\sigma_q(\tau_i)-\tau_i|<\frac18 f(m_i)$ and such that $|\sigma_q(\tau_i)-\tau_i|<|\sigma_n(\tau_i)-\tau_i|$ for all $n<q$ (notice that $\sigma_n(\tau_i)\neq\tau_i$ for all n since $\tau_i\notin\mathbb{Q}$). Notice that this last inequality implies that $\sigma_q(\tau_i)\in I_q(\tau_i)$. It follows that $I_{m_i}(\tau_i)\cap I_q(\tau_i)$ is an interval of non-zero measure containing $\sigma_q(\tau_i)$, and so we can find an irrational $\tau_{i+1}\in I_{m_i}(\tau_i)\cap I_q(\tau_i)$

with $|\tau_{i+1}-\sigma_q(\tau_i)|<\frac{1}{2}f(q)$. Then, defining $m_{i+1}:=q$, we have that τ_{i+1} and m_{i+1} have the required properties since one has $\tau_{i+1},\tau_i\in I_{m_i}(\tau_i)$ and

$$|\tau_{i+1}-\sigma_{m_{i+1}}(\tau_{i+1})|=|\tau_{i+1}-\sigma_{m_{i+1}}(\tau_i)|<\tfrac{1}{2}f(m_{i+1})$$
$$|\tau_{i+1}-\tau_i|\le|\sigma_q(\tau_i)-\tau_i|+|\sigma_q(\tau_i)-\tau_{i+1}|<\tfrac{1}{8}f(m_i)+\tfrac{1}{2}f(m_{i+1})\le\tfrac{1}{4}f(m_i),$$

where the equality on the first line follows since $\tau_{i+1}\in I_{m_{i+1}}(\tau_i)$. □

6. The Thue–Morse constant

Let the family of sequences $\mathcal{E}_k:=(\mathcal{E}_k(n))_{n\ge0}$ be defined recursively by

$$\mathcal{E}_0(n):=\varepsilon_n\quad\text{and}\quad\mathcal{E}_{k+1}(n):=\sum_{m\le n}\mathcal{E}_k(m)\quad\text{for }k,n\ge0,$$

so that the k-th sequence is the sequence of partial sums of the $(k-1)$-th sequence, the first one being the Thue–Morse sequence. The next lemma gives a description of the sequences $\mathcal{E}_k$.

Proposition 6.1. *For every $k\ge0$ there exists a finite sequence $W_k:=(w_k(n))_{n=0}^{2^k-1}$, with $w_k(n)\in\mathbb{Z}_{\ge0}$, such that:*

$$\mathcal{E}_k=(\varepsilon_0\cdot W_k,\varepsilon_1\cdot W_k,\varepsilon_2\cdot W_k,\dots).\tag{6.1}$$

Moreover, one has:

(*i*) $\sum_{n=0}^{2^k-1}w_k(n)=2^{\binom{k}{2}}$;

(*ii*) $w_k(2^k-1)=\cdots=w_k(2^k-k)=0$;

(*iii*) $w_k(2^k-k-1-n)=w_k(n)$ *for all n, with* $0\le n\le2^k-k-1$.

(*iv*) $w_k(n)\le2^{\binom{k-1}{2}}$ *for all $n\in[0,2^k)$ and the equality holds if and only if* $2^{k-1}-k-1\le n\le 2^{k-1}-1$.

Proof. We give a proof using generating functions. Given a sequence of integers $(a_n)_{n\ge0}$ with generating function $F(x)$, the generating function of $\left(\sum_{m\le n}a_n\right)_{n\ge0}$ is equal to $F(x)/(1-x)$. Now, letting $E_k(x)$ denote the generating function of $\mathcal{E}_k$ for $k\ge0$, we have

$$E_0(x):=\sum_{n=0}^{\infty}\varepsilon_nx^n=\prod_{j=0}^{\infty}(1-x^{2^j})$$

and thus

$$E_k(x)=\frac{E_0(x)}{(1-x)^k}=\prod_{j=0}^{k-1}\frac{1-x^{2^j}}{1-x}\prod_{j=k}^{\infty}(1-x^{2^j})=Q_k(x)E_0(x^{2^k}),\tag{6.2}$$

where

$$Q_k(x):=\prod_{j=0}^{k-1}(1+x+x^2+\cdots+x^{2^j-1})=\sum_{n=0}^{2^k-1}w_k(n)x^n,\tag{6.3}$$

for some sequence of integers $(w_k(n))_{n=0}^{2^k-1}$ with

$$w_k(n):=\#\{(a_0,\dots,a_{k-1})\in\mathbb{Z}_{\ge0}^k\colon a_0+\cdots+a_{k-1}=n,\ a_j\le2^j-1\ \forall j\}.\tag{6.4}$$

At this point, (6.1) follows immediately from (6.2). Also, $Q_k(1)=2^{\binom{k}{2}}$ implies (*i*), and $\deg Q_k=2^k-k-1$ implies (*ii*). Furthermore, we have $Q_k(x)=Q_k(1/x)x^{2^k-k-1}$ and so (*iii*) follows. Finally, if n satisfies

$$\sum_{j=0}^{k-2}(2^j-1)=2^{k-1}-k-1\le n\le2^{k-1}-1,$$

then for any $a_0, \dots, a_{k-2}$ as in (6.4) there exists a unique a_{k-1} such that $a_0 + \cdots + a_{k-1} = n$. Thus, for all such n one has $w_k(n) = \prod_{j=0}^{k-2} 2^j = 2^{\binom{k-1}{2}}$ and the same reasoning gives $w_k(n) < 2^{\binom{k-1}{2}}$ for all the other possible values of n, so that (iv) follows. □

Corollary 6.2. *For $n \geq 1$ let $n = 2^\mu n'$ with n' odd. Then $\mathcal{E}_1(n-1) = \mathcal{E}_2(n-1) = \cdots = \mathcal{E}_\mu(n-1) = 0$ and $\mathcal{E}_{\mu+1}(n-1) = -2^{\binom{\mu}{2}}\varepsilon_n$.*

Proof. By (6.1) we have $|\mathcal{E}_k(n-1)| = |\mathcal{E}_k(m-1)|$ if $m \equiv n \pmod{2^k}$. In particular, if $\mu \leq k$, then $|\mathcal{E}_k(n-1)| = |\mathcal{E}_k(2^k-1)| = 0$ by Proposition 6.1, (ii). Moreover writing $n' = 2n'' + 1$ we have $n - 1 = 2^{\mu+1}n'' + 2^\mu - 1$. Thus, by (6.1) and Proposition 6.1, (iv), we have $\mathcal{E}_{\mu+1}(n-1) = 2^{\binom{\mu}{2}}\varepsilon_{n''} = -2^{\binom{\mu}{2}}\varepsilon_{n'} = -2^{\binom{\mu}{2}}\varepsilon_n$ (because the multiplication by 2 does not modify the number of non-zero digits in the binary representation). □

The first few sequences W_k are as follows:

$$W_0 = (1), \qquad W_1 = (1,0), \qquad W_2 = (1,1,0,0),$$
$$W_3 = (1,2,2,2,1,0,0,0), \qquad W_4 = (1,3,5,7,8,8,8,8,7,5,3,1,0,0,0,0),$$
$$W_5 = (1,4,9,16,24,32,40,48,55,60,63,64,64,64,64,64,63,\dots,1,0,\dots,0).$$

Remark 6.3. The sequence arising from the W_k also appeared and was studied in the recent work [33, Section 5]. In this paper Vignat and Wakhare defined the numbers α_m^N such that

$$\sum_{m_1=0}^{2^1-1} \cdots \sum_{m_N=0}^{2^N-1} g(m_1 + \cdots + m_N) = \sum_{m=0}^{2^N-N-1} \alpha_m^N g(m).$$

for any function g. It is not difficult to see (cf. (6.3) and [33, (5.1)]) that the two sequences coincide, i.e., $w_{k-1}(n) = \alpha_n^k$ for $k \geq 1$, $0 \leq n < 2^k$.

Remark 6.4. The vectors W_k are related to the Fabius function. This can be defined as the unique solution $F : [0,1] \to \mathbb{R}$ of the following functional differential equation problem

$$\begin{aligned} F(0) &= 0; \\ F(1-x) &= 1 - F(x), \quad x \in [0,1]; \\ F'(x) &= 2F(2x), \quad x \in [0,1/2]. \end{aligned}$$

It can also be defined as the cumulative distribution function of $\sum_{n=1}^\infty X_n/2^n$, where $X_1, X_2, \dots$ are independent and uniformly distributed random variables on the unit interval $[0,1]$. It is an example of an infinitely differentiable function that is nowhere analytic [17, 20]. Setting $F_k'(n/2^k) := 2^{(3k-k^2)/2} w_k(n)$, for all integers $k \geq 0$ and $n \in [0, 2^k)$, we have the identities

$$\begin{aligned} &\frac{1}{2^k}\sum_{n=0}^{2^k-1} F_k'(n/2^k) = 1; \\ &F_k'\Big(1 - \frac{n+k+1}{2^k}\Big) = F_k'\left(\frac{n}{2^k}\right), \qquad n \in [0, \dots, 2^k - k - 1]; \\ &F_k'\Big(\frac{n}{2^k}\Big) = 2 \cdot \frac{1}{2^{k-1}} \sum_{m \leq n} F_{k-1}'\Big(\frac{m}{2^{k-1}}\Big), \quad n \in [0, \dots, 2^{k-1} - 1]; \end{aligned}$$

which are discrete versions of

$$\begin{aligned} &\int_0^1 F'(x)\,\mathrm{d}x = 1; \\ &F'(1-x) = F'(x), \quad x \in [0,1]; \\ &F'(x) = 2\int_0^{2x} F'(t)\,\mathrm{d}t, \quad x \in [0,1/2]. \end{aligned}$$

Indeed, $F_k'(x)$ approximates $F'(x) = 2F(2x)$ as $k \to +\infty$ (see Figure 3). See [8, 9, 22] for more information about the values of the Fabius function at dyadic fractions.

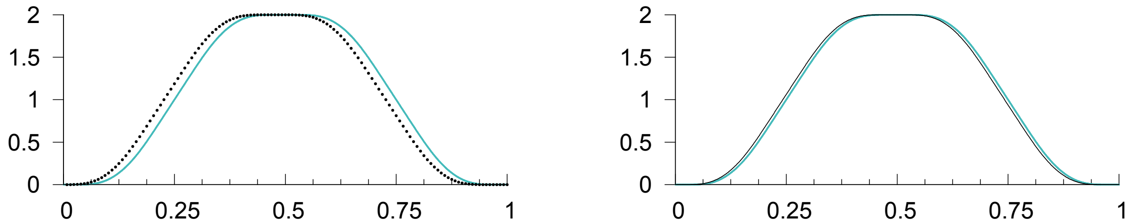

FIGURE 3. The graphs of the derivative of the Fabius function (solid lines) and of the vectors W_6 (left, dotted) and W_9 (right, dotted). The vectors are scaled to fit in $[0, 2]$.

Proposition 6.1 shows that the sequence of partial sums $\mathcal{E}_1$ is bounded. In particular, by partial summation the series

$$\tau_0 := \sum_{n=1}^{+\infty} \frac{\varepsilon_{n-1}}{n} = \sum_{n=1}^{+\infty} \mathcal{E}_1(n-1)\left(\frac{1}{n} - \frac{1}{n+1}\right) \tag{6.5}$$

converges. Applying partial summation repeatedly, one also obtains the identity

$$\tau_0 - \sum_{m=1}^{n} \frac{\varepsilon_{m-1}}{m} = \sum_{m>n} \frac{\varepsilon_{m-1}}{m} = -\sum_{\ell=1}^{k} \mathcal{E}_\ell(n-1)G_{\ell-1}(n+1) + \sum_{m>n} \mathcal{E}_k(m-1)G_k(m), \tag{6.6}$$

for all $k \geq 0$ and $n \geq 1$, where G_k is defined by the recurrence relation

$$G_0(x) := \frac{1}{x}, \quad \text{and} \quad G_{k+1}(x) := G_k(x) - G_k(x+1) \qquad \text{for } k \geq 1.$$

Notice that the definition is similar to the one of g_k, but the shift in the second argument is different. The following lemma shows that G_k can be written also as a simple rational function.

Lemma 6.5. *For every k we have*

$$G_k(x) = \frac{k!}{\prod_{\ell=0}^{k}(x+\ell)}.$$

In particular, $0 < G_k(x) \leq k!/x^{k+1}$ when $x > 0$ and $G_k(x)x^{k+1} = k!(1 + O(1/x))$ as $x \to \infty$.

Proof. We proceed by induction on k. The formula is trivial for $k = 0$. Using the recursive definition of G_k and the inductive hypothesis we get:

$$\begin{aligned} G_{k+1}(x) = G_k(x) - G_k(x+1) &= \frac{k!}{\prod_{\ell=0}^{k}(x+\ell)} - \frac{k!}{\prod_{\ell=0}^{k}(x+\ell+1)} \\ &= \frac{k!}{\prod_{\ell=0}^{k}(x+\ell)} - \frac{k!}{\prod_{\ell=1}^{k+1}(x+\ell)} = \frac{(k+1)!}{\prod_{\ell=0}^{k+1}(x+\ell)}. \end{aligned}$$

□

We now give the following lemma which implies that the series defining τ_0 gives its greedy representation.

Lemma 6.6. *We have $\tau_0 > 0$. Moreover, for $n \geq 1$ let $n = 2^\mu n'$ with n' odd. Then*

$$\varepsilon_n \sum_{m>n}^{\infty} \frac{\varepsilon_{m-1}}{m} > 2^{\binom{\mu}{2}} G_{\mu+1}(n + 2^{\mu+1}) > 0. \tag{6.7}$$

Proof. Since $|\mathcal{E}_1(n)| \leq 1$ for all n and $\mathcal{E}_1(0) = 1$, $\mathcal{E}_1(1) = 0$, by (6.5) we have

$$\tau_0 \geq \frac{1}{2} - \sum_{n=3}^{+\infty} \Big(\frac{1}{n} - \frac{1}{n+1}\Big) = \frac{1}{6}.$$

In particular $\tau_0 > 0$.

Applying (6.6) with $k = \mu + 1$ gives

$$\sum_{m>n} \frac{\varepsilon_{m-1}}{m} = -\,\mathcal{E}_{\mu+1}(n-1)G_\mu(n+1) + \sum_{m>n} \mathcal{E}_{\mu+1}(m-1)G_{\mu+1}(m), \tag{6.8}$$

since by Corollary 6.2 one has $\mathcal{E}_\ell(n-1) = 0$ for all $\ell \leq \mu$. By Proposition 6.1, (iv), and the periodicity of $(|\mathcal{E}_{\mu+1}(n)|)_n$, we have $|\mathcal{E}_{\mu+1}(m)| \leq 2^{\binom{\mu}{2}}$ and $\mathcal{E}_{\mu+1}(c-1) = 0$ where c is the smallest multiple of $2^{\mu+1}$ such that $c > n$; in particular, $c < n + 2^{\mu+1}$. Thus,

$$\Big|\sum_{m>n} \mathcal{E}_{\mu+1}(m-1)G_{\mu+1}(m)\Big| \leq 2^{\binom{\mu}{2}}\Big(\sum_{m>n} G_{\mu+1}(m) - G_{\mu+1}(c)\Big) = 2^{\binom{\mu}{2}}\big(G_\mu(n+1) - G_{\mu+1}(c)\big),$$

where in the last two steps we have used the positivity and the telescoping property of $G_{\mu+1}$. Also, by Corollary 6.2 we have $\mathcal{E}_{\mu+1}(n-1) = -\varepsilon_n\, 2^{\binom{\mu}{2}}$. Thus, (6.8) gives

$$2^{\binom{\mu}{2}}G_\mu(n+1) - \varepsilon_n \sum_{m>n} \frac{\varepsilon_{m-1}}{m} < \Big|\sum_{m>n} \mathcal{E}_{\mu+1}(m-1)G_k(m)\Big| < 2^{\binom{\mu}{2}}\big(G_\mu(n+1) - G_{\mu+1}(c)\big)$$

and the claimed inequality follows since $G_{\mu+1}(c) \geq G_{\mu+1}(n + 2^{\mu+1})$. □

We are now in position to prove the claims in Theorem 1.9.

Proof of Theorem 1.9. The fact that $s_n = s_n(\tau_0) = \varepsilon_{n-1}$ for all $n \geq 1$ follows from Lemma 6.6, by proceeding as in the proof of Corollary 4.4.

Let $k \geq 1$ and let $n = 2^k n'$ with n' odd. Corollary 6.2 shows that $\mathcal{E}_1(n-1) = \cdots = \mathcal{E}_k(n-1) = 0$. Also, Proposition 6.1 gives that $|\mathcal{E}_k(m)| \leq 2^{\binom{k-1}{2}}$ for all m. Hence, (6.6) gives

$$|\tau_0 - \sigma_n| = \Big|\sum_{m>n} \mathcal{E}_k(m-1)G_k(m)\Big| \leq 2^{\binom{k-1}{2}} \sum_{m>n} G_k(m),$$

since $G_k(m)$ is positive. Recalling the telescoping definition of G_k and the bound for G_{k-1} given in Lemma 6.5, we get

$$|\tau_0 - \sigma_n| \leq 2^{\binom{k-1}{2}}G_{k-1}(n+1) \leq 2^{\binom{k-1}{2}}(k-1)!/n^k, \tag{6.9}$$

which is the first claim, since $2^{\binom{k-1}{2}}(k-1)! = c_{k-1}$. Moreover, applying (6.6) with $k+2$ in place of k, gives

$$\tau_0 - \sigma_n = -\,\mathcal{E}_{k+1}(n-1)G_k(n+1) - \mathcal{E}_{k+2}(n-1)G_{k+1}(n+1) + \sum_{m>n} \mathcal{E}_{k+2}(m-1)G_{k+2}(m).$$

Corollary 6.2 shows that $\mathcal{E}_{k+1}(n-1) = -2^{\binom{k}{2}}\varepsilon_n$, whereas, telescoping, again, we have

$$\Big|\sum_{m>n} \mathcal{E}_{k+2}(m-1)G_{k+2}(m)\Big| < 2^{\binom{k+1}{2}}G_{k+1}(n+1) = O_k(n^{-k-2})$$

Thus,

$$\tau_0 - \sigma_n = \frac{\varepsilon_n c_k}{n^{k+1}}(1 + O_k(1/n)) \sim_k \frac{\varepsilon_n c_k}{n^{k+1}}.$$

Finally, by (6.7) we have

$$|\tau_0 - \sigma_n| > 2^{\binom{k}{2}}G_{k+1}(n + 2^{k+1}) > 2^{\binom{k}{2}}G_{k+1}(3n) > 2^{\binom{k}{2}-2k-2}\frac{(k+1)!}{n^{k+2}}$$

since $n \geq 2^k \geq (k+1)$. Thus, after a quick computation with Stirling's formula one obtains

$$\log|\tau_0 - \sigma_n| > \min_{k\in\mathbb{R}} \Big(\frac{\log 2}{2}(k^2-5k) + \log(k!) - 2 - k\log n\Big) = -\frac{(\log n - \log\log n)^2}{\log 4} + O(\log n).$$

Equation (1.10) then follows since (6.9) with $n = 2^k$ gives

$$|\tau_0 - \sigma_n| \leq 2^{-\frac{k^2+3k}{2}+1}(k-1)! = e^{-\frac{(\log n - \log\log n)^2}{\log 4} + O(\log n)}. \qquad \square$$

References

1. J.-P. Allouche, *Thue, combinatorics on words, and conjectures inspired by the Thue-Morse sequence*, J. Théor. Nombres Bordeaux **27** (2015), no. 2, 375–388. MR 3393159
2. J.-P. Allouche and H. Cohen, *Dirichlet series and curious infinite products*, Bull. London Math. Soc. **17** (1985), no. 6, 531–538. MR 0813734
3. J.-P. Allouche, H. Cohen, M. Mendès France, and J. O. Shallit, *De nouveaux curieux produits infinis*, Acta Arith. **49** (1987), no. 2, 141–153. MR 928633
4. J.-P. Allouche, S. Riasat, and J. Shallit, *More infinite products: Thue–Morse and the gamma function*, Ramanujan J. **49** (2019), no. 1, 115–128. MR 3942297
5. J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge University Press, Cambridge, 2003, Theory, applications, generalizations. MR 1997038
6. J.-P. Allouche and J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence*, Sequences and their applications (Singapore, 1998), Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999, pp. 1–16. MR 1843077
7. J.-P. Allouche and J. O. Shallit, *Infinite products associated with counting blocks in binary strings*, J. London Math. Soc. (2) **39** (1989), no. 2, 193–204. MR 0991655
8. J. Arias de Reyna, *Definición y estudio de una función indefinidamente diferenciable de soporte compacto*, Rev. Real Acad. Cienc. Exact. Fís. Natur. Madrid 76 (1982), no. 1, 21–38 (Spanish). English version: `https://arxiv.org/abs/1702.05442`.
9. J. Arias de Reyna, *Arithmetic of the Fabius function*, (2017), `https://arxiv.org/abs/1702.06487`.
10. A. G. Beged-dov, *Lower and upper bounds for the number of lattice points in a simplex*, SIAM J. Appl. Math. **22** (1972), 106–108. MR 0307685
11. S. Bettin, G. Molteni, and C. Sanna, *Small values of signed harmonic sums*, C. R. Math. Acad. Sci. Paris **356** (2018), no. 11-12, 1062–1074. MR 3907571
12. M. Dekking, *On the structure of Thue-Morse subwords, with an application to dynamical systems*, Theoret. Comput. Sci. **550** (2014), 107–112. MR 3248841
13. M. Dekking, M. Mendès France, and A. van der Poorten, *Folds!*, Math. Intelligencer **4** (1982), no. 3, 130–138. MR 0684028
14. M. Dekking, M. Mendès France, and A. van der Poorten, *Folds! II. Symmetry disturbed*, Math. Intelligencer **4** (1982), no. 4, 173–181. MR 0685557
15. M. Dekking, M. Mendès France, and A. van der Poorten, *Folds! III. More morphisms*, Math. Intelligencer **4** (1982), no. 4, 190–195. MR 0685559
16. M. Dekking, M. Mendès France, and A. van der Poorten, *Letters to the editor*, Math. Intelligencer **5** (1983), no. 5. (It contains a corrigendum to [14, Observation 2.13].)
17. Y. Dimitrov and G. A. Edgar, *Solutions of self-differential functional equations*, Real Anal. Exchange **32** (2006/07), no. 1, 29–54. MR 2329220
18. S. Dubuc and A. Elqortobi, *Le maximum de la fonction de Knopp*, INFOR Inf. Syst. Oper. Res. **28** (1990), 311–323.
19. M. Euwe, *Mengentheoretische Betrachtungen über das Schachspiel*, Proc. Akad. Wet. Amsterdam **32** (1929), 633–642.
20. J. Fabius, *A probabilistic example of a nowhere analytic C^∞-function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **5** (1966), 173–174. MR 0197656
21. R. K. Guy, *Unsolved problems in number theory*, third ed., Problem Books in Mathematics, Springer-Verlag, New York, 2004. MR 2076335
22. J. K. Haugland, *Evaluating the Fabius function*, (2016), `https://arxiv.org/abs/1609.07999`.
23. V. M. Kadets and M. I. Kadets, *Rearrangements of series in Banach spaces*, Translations of Mathematical Monographs, vol. 86, American Mathematical Society, Providence, RI, 1991, Translated from the Russian by Harold H. McFaden. MR 1108619
24. V. Komornik and P. Loreti, *Unique developments in non-integer bases*, Amer. Math. Monthly **105** (1998), no. 7, 636–639. MR 1633077

25. M. Lothaire, *Combinatorics on words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1997, With a foreword by Roger Lyndon and a preface by Dominique Perrin, Corrected reprint of the 1983 original, with a new preface by Perrin. MR 1475463
26. C. Mauduit and J. Rivat, *La somme des chiffres des carrés*, Acta Math. **203** (2009), no. 1, 107–148. MR 2545827
27. C. Mauduit and J. Rivat, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. of Math. (2) **171** (2010), no. 3, 1591–1646. MR 2680394
28. H. M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. **22** (1921), no. 1, 84–100. MR 1501161
29. E. Prouhet, *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sci. Paris Sér. I Math. **33** (1851), 225.
30. D. Robbins, *Problems and Solutions: Solutions of Elementary Problems: E2692*, Amer. Math. Monthly **86** (1979), no. 5, 394–395. MR 1539042
31. J. B. Rosser and L. Schoenfeld, *Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$*, Math. Comp. **29** (1975), 243–269. MR 0457373
32. J. O. Shallit, *On infinite products associated with sums of digits*, J. Number Theory **21** (1985), no. 2, 128–134. MR 0808281
33. C. Vignat and T. Wakhare, *Finite generating functions for the sum of digits sequence*, Ramanujan J. (2018), `https://doi.org/10.1007/s11139-018-0065-0`
34. L. C. Washington, *Introduction to cyclotomic fields*, second ed., Springer-Verlag, New York, 1997. MR 1421575
35. D. R. Woods, *Problems and Solutions: Elementary Problems: E2692*, Amer. Math. Monthly **85** (1978), no. 1, 48. MR 1538582

(S. Bettin) Dipartimento di Matematica, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy

Email address: `bettin@dima.unige.it`

(G. Molteni) Dipartimento di Matematica, Università di Milano, Via Saldini 50, 20133 Milano, Italy

Email address: `giuseppe.molteni1@unimi.it`

(C. Sanna) Dipartimento di Matematica, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy

Email address: `carlo.sanna.dev@gmail.com`